Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

"Novel and rapid tests for diagnosis of tuberculosis using non-sputum specimens"

by
Stephanie Minnies

Dissertation presented for the degree of
Doctor of Philosophy (Molecular Biology) in the Faculty of Medicine and Health Sciences at Stellenbosch University

Supervisor: Prof Grant Theron
Co-supervisor: Dr Byron Reeve

December 2023

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification. This work was conducted within the Centre for Tuberculosis Research of the South African Medical Research Council.

December 2023

Acknowledgements

I would firstly like to thank my supervisor, Prof. Grant Theron, for undoubtedly believing I would complete this journey when I was uncertain, for his sometimes longwinded but always helpful explanations, for even agreeing to have a meeting while he was in the middle of Albania with limited service to help me when I needed it. Your guidance has been invaluable.

To my co-supervisor, Dr. Byron Reeve, thank you for your constant support, your patience and always being willing to help or be a guiding force. I always felt that I was lucky to have you as a supervisor, and I will never forget your kindness and faith in me.

I want to thank Sr Ruth Wilson, who was instrumental in recruiting patients to the study. Her ability to talk to you, give you advice, recruit 10 patients a day and still tell you how you can do something better made such an impact on me. You were a whirlwind Ruth, and you inspired me to always be better.

Thank you to everyone in CLIME lab for always offering to help when needed, and providing humour to situations that would normally cause tears. Thank you to Dr. Rouxjeane Venter and Dr. Brigitta Derendinger for helping and guiding me when I was the new minion to join CLIME lab. Thank you to Zaida Palmer, who I could turn to when I needed help adjusting to an unexpected aspect of research. And finally, thank you to Sr. Dawn Davis for giving a sweet and encouraging word when I needed one. These kindnesses will not be forgotten.

I would like to acknowledge the funding and support of the South African Medical Research Council (SAMRC), the National Research Foundation (NRF), and Harry Crossley Foundation (HCF). Opinions expressed and conclusions arrived do not represent the views of the SAMRC, NRF or HCF.

Thank you to Taylor Swift for writing Folklore, Speak Now and Red, and Hayley Williams for writing Petals for Armor. Those albums helped me get through many, many writing sessions.

Thank you to my dear friend, Stephanie Kala. Your love, support, and the joy you bring to my life mean the world to me.

Thank you, Uncle, for always encouraging me to continue while also telling me it's time to stop studying now. Thank you, Ousie, for always being a voice of reason in my life and making those countless birthday cakes from scratch. Thank you, aunty Maverine, your overabundance of confidence never failed to make me laugh and bring joy to my life. You are all missed but will never be forgotten.

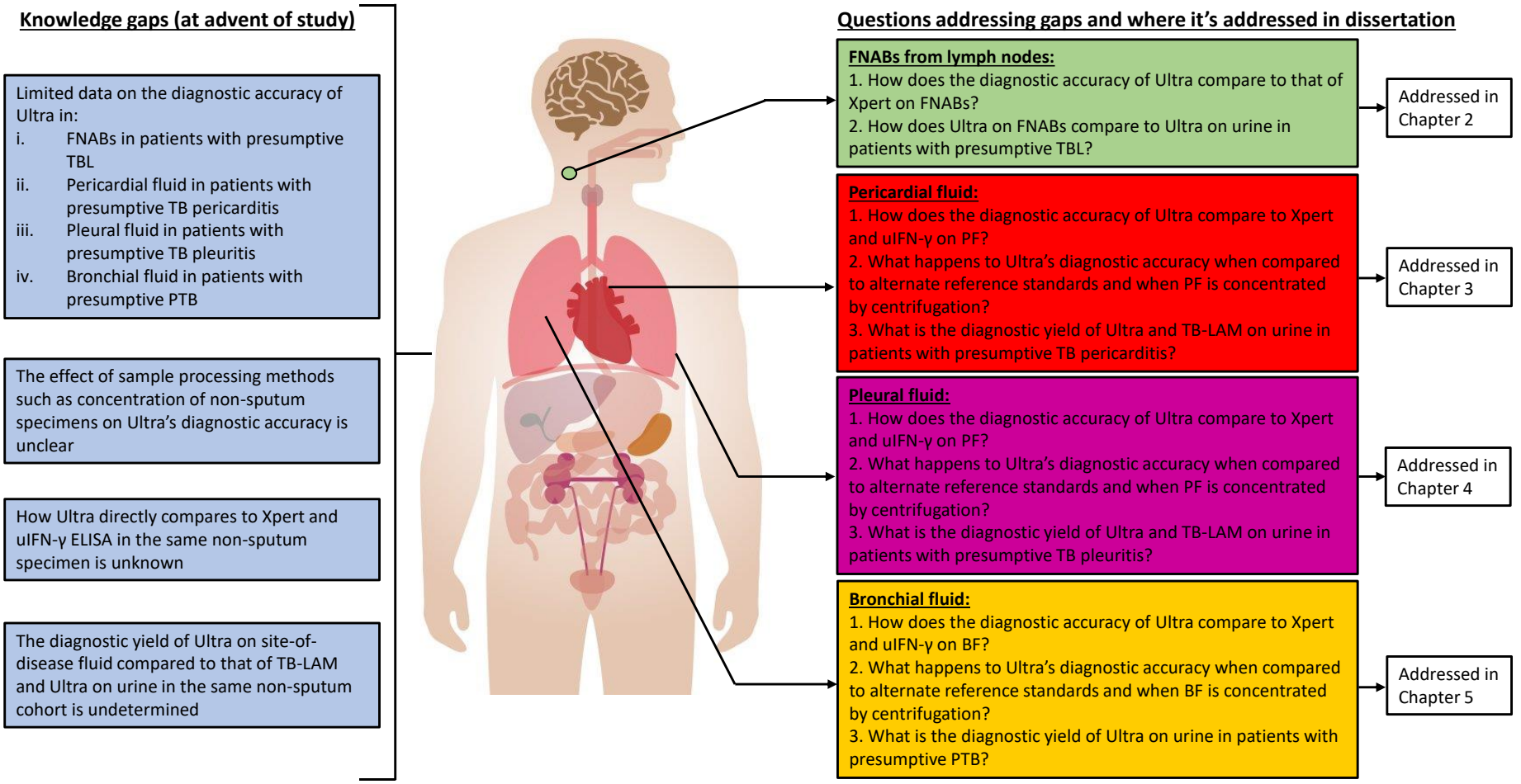
Thank you to my main ally in this world, my sister Merlin. I'm so beyond grateful to have you in my life. You are the baby, and you know how to use that to your advantage, but when I need it, you easily and gladly step up. To make me laugh, be the shoulder I can cry on, allow me to vent about things you have no clue about. Thank you for being my biggest cheerleader. I could not have done this without you.

Thank you, Dad, for always trusting me to know what's best for me. Thank you for always being there for me, and for being my champion when I needed it. Mom, how I wish you could be here to see this. Your humour, sassiness and strength has stayed with me, and has helped me through this journey so many times. Thank you for being the best mother I could have asked for. You showed me how to forge ahead, and I always will because you taught me how to.

"Of course, it's happening inside your head Harry, why should that mean that it's not real?"

- Albus Dumbledore

Graphical abstract: Overview of the knowledge gaps in the diagnosis of tuberculosis using non-sputum specimens and how this dissertation aims to address these gaps



Abbreviations: BF, bronchial fluid; TB-LAM, Determine TB-LAM; FNABs, fine needle aspirate biopsies; PF, pericardial fluid (chapter 3)/pleural fluid (chapter 4); PTB, pulmonary TB; TBL, TB lymphadenitis; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Summary

Diagnosis of TB remains a challenge, as in 2021, 60% of those who developed active TB were diagnosed. Xpert Ultra MTB/RIF (Ultra) is endorsed for TB diagnosis on sputum, but at the advent of the study, more data on the usefulness of Ultra on non-sputum specimens, particularly in HIV-endemic settings, were needed. Moreover, the impact of different sample processing such as sample concentration of non-sputum specimens by centrifugation on Ultra has not been previously investigated. Lastly, data on how Ultra on site-of-disease non-sputum specimens directly compares to other tests, either on site-of-disease or non-site-of-disease non-sputum specimens remain limited.

Firstly (chapter 2), we showed that in patients with presumptive TB lymphadenitis (TBL), Ultra detects more TBL cases than Xpert MTB/RIF (Xpert), Ultra's predecessor, and results in more people being placed on treatment. Ultra's increased sensitivity on fine needle aspirate biopsies (FNABs) does however come with decreased specificity, and this was not significantly associated with HIV status or the use of alternate reference standards. Furthermore, we showed that study Ultra detected more TBL cases than programmatic Ultras when both tests were done, indicating that optimisation of programmatic testing of FNABs would result in improved TBL diagnosis. Moreover, we showed that FNAB Ultra false-negative results are associated with PCR inhibition. Lastly, we showed that in patients with presumptive TBL, urine-Ultra had low sensitivity.

Thereafter (chapter 3), we found that in people living with HIV (PLHIV) with presumptive TB pericarditis, Ultra on unconcentrated pericardial fluid had higher sensitivity and lower specificity overall when compared to Xpert. We also found that comparing Ultra to alternate reference standards did not improve sensitivity. Exclusion of Ultra results is the superior recategorization strategy in pericardial fluid (unlike reclassifying trace results as negative).

Additionally, we showed that using concentrated pericardial fluid on Ultra resulted in higher Ultra specificity but more non-actionable results. This suggests that laboratories with adequate fluid volume and capacity should concentrate pericardial fluid when possible. Furthermore, we showed that the high sensitivity of uIFN-γ on pericardial fluid is offset by its' poor specificity, indicating that Ultra is the superior test on pericardial fluid. Lastly, Urinary Ultra and TB-LAM had low sensitivity but could reduce the need for pericardiocentesis for TB pericarditis diagnosis in 4% of patients, highlighting their potential.

Thirdly (chapter 4), in patients with presumptive TB pleuritis, we showed that Ultra had similar sensitivity but higher diagnostic yield compared to Xpert, and exclusion or reclassification of trace results to negative does not increase sensitivity. Additionally, alternate reference standards and HIV status did not significantly increase Ultra's sensitivity. Furthermore, we showed that testing Ultra with concentrated pleural fluid increases Ultra specificity, but this also increases non-actionable results, and this was also observed in pericardial fluid. Moreover, we showed that uIFN-γ on pleural fluid had high sensitivity and moderate specificity, suggesting that laboratories with sufficient funding and infrastructure should use uIFN-γ concentration for TB pleuritis diagnosis. Finally, we showed that Ultra and TB-LAM on urine could reduce the need for thoracentesis in a subset of patients for TB pleuritis diagnosis, particularly in PLHIV.

Lastly (chapter 5), we showed that in bronchial fluid (BF), Ultra's diagnostic accuracy was not significantly different between bronchoalveolar lavage fluid (BALF) and bronchial wash fluid (BWF), and thus they were not stratified in downstream analyses for TB diagnosis. We also showed that Ultra on concentrated BF had higher sensitivity and lower specificity when compared to Xpert (HIV and alternate reference standards did not significantly change Ultra's sensitivity and specificity). Moreover, 4 in 5 Ultra "false-positives" started empirical treatment, which suggests that Ultra on BF could be detecting TB cases missed by culture. We also

showed that programmatic Ultra testing on BF would benefit from optimisation as study Ultras detected more TB cases. Moreover, we showed that uIFN-γ should not be used on BF for TB diagnosis due to its' poor sensitivity. Lastly, we showed that urinary-Ultra had low sensitivity, but still detected TB missed by tests on site-of-disease fluid, highlighting its' usefulness.

In terms of outputs, this dissertation has resulted in four first author manuscripts. One has been published in a peer reviewed journal (chapter 2) and the others' (chapters 3, 4 and 5) are submission ready. Additionally, three ancillary publications (one of which was co-first authored) are briefly discussed in chapter 8 and can be found in the appendices. Some of this research was presented by the candidate at an international and national peer-reviewed conference.

In summary, this work shows Ultra's high sensitivity and moderate sensitivity on FNABs, pericardial fluid, pleural fluid, and BF in patients with presumptive TBL, TB pericarditis, TB pleuritis and PTB. We can therefore recommend a positive Ultra, with the inclusion of trace results, for TB diagnosis in these non-sputum specimens.

Opsomming

Diagnose van TB bly 'n uitdaging, want in 2021 is 60% van diegene wat aktiewe TB ontwikkel het is gediagnoseer. Xpert Ultra MTB/RIF (Ultra) word onderskryf vir TB-diagnose op sputum, maar met die koms van die studie was meer data oor die bruikbaarheid van Ultra op nie-sputummonsters, veral in MIV-endemiese omgewings, nodig. Boonop is die impak van verskillende monsterverwerking soos monsterkonsentrasie van nie-sputummonsters deur sentrifugering op Ultra nie voorheen ondersoek nie. Laastens, data oor hoe Ultra op plek-van-siekte nie-sputummonsters direk kan vergelyk word met ander toetse op plek-van-siekte of nie-plek-van-siekte nie-sputummonsters bly steeds beperk.

Eerstens (hoofstuk 2) het ons getoon dat in pasiënte met vermoedelike TB limfadenitis (TBL), Ultra meer TBL-gevalle opgespoor as Xpert MTB/RIF (Xpert), Ultra se voorganger, en dit het aanleiding gegee dat meer mense op behandeling geplaas word. Ultra se verhoogde sensitiwiteit op fynnaald aspiraatbiopsies (FNABs) kom egter met verminderde spesifisiteit, en dit was nie beduidend geassosieer met MIV-status of die gebruik van alternatiewe verwysingstandaarde nie. Verder het ons getoon dat studie Ultra meer TBL-gevalle as programmatiese Ultras opgespoor het wanneer beide toetse gedoen is, wat aandui dat optimalisering van programmatiese toetsing van FNABs verbeterde TBL-diagnose tot gevolg sou hê. Verder het ons getoon dat FNAB Ultra vals-negatiewe resultate geassosieer word met PCR-inhibisie. Laastens het ons getoon dat by pasiënte met vermoedelike TBL, urine-Ultra lae sensitiwiteit tot gevolg gehad het.

Daarna (hoofstuk 3) het ons gevind dat in mense met MIV (MMIV) en met vermoedelike TB-perikarditis, Ultra op ongekonsentreerde perikardiale vloeistof hoër sensitiwiteit en laer spesifisiteit in die algemeen gehad het in vergelyking met Xpert. Ons het ook gevind dat die vergelyking van Ultra met alternatiewe verwysingstandaarde nie sensitiwiteit verbeter het nie.

Uitsluiting van Ultra-resultate is die beter herkategoriseringstrategie in perikardiale vloeistof (anders as om spoor as negatief te herklassifiseer). Daarbenewens het ons getoon dat die gebruik van gekonsentreerde perikardiale vloeistof op Ultra gelei het tot hoër Ultra spesifisiteit, maar meer nie-uitvoerbare resultate. Dit dui daarop dat laboratoriums met voldoende vloeistofvolume en kapasiteit, waar moontlik, perikardiale vloeistof moet konsentreer. Verder het ons getoon dat die hoë sensitiwiteit van uIFN-γ op perikardiale vloeistof geneutraliseer word deur sy swak spesifisiteit, wat aandui dat Ultra die superieure toets op perikardiale vloeistof is. Laastens, Urinary Ultra en TB-LAM het lae sensitiwiteit gehad, maar kon die behoefte aan perikardiosentese vir TB-perikarditis-diagnose in 4% van pasiënte verminder, wat hul potensiaal beklemtoon het.

Derdens (hoofstuk 4), in pasiënte met vermoedelike TB-pleuritis, het ons getoon dat Ultra soortgelyke sensitiwiteit maar hoër diagnostiese opbrengs in vergelyking met Xpert gehad het, en uitsluiting of herklassifikasie van spoorresultate na negatief en nie sensitiwiteit verhoog nie. Daarbenewens het alternatiewe verwysingstandaarde en MIV-status nie Ultra se sensitiwiteit aansienlik verhoog nie. Verder het ons getoon dat die toets van Ultra met gekonsentreerde pleurale vloeistof Ultra-spesifisiteit verhoog, maar dit verhoog ook nie-werkbare resultate, en dit is ook waargeneem in perikardiale vloeistof. Verder het ons getoon dat uIFN-γ op pleurale vloeistof hoë sensitiwiteit en spesifisiteit het, wat daarop dui dat laboratoriums met die nodige befondsing en infrastruktuur die toets moet gebruik vir TB pleuritis diagnose. Ten slotte het ons getoon dat Ultra en TB-LAM die behoefte aan torasentese in 'n subset van pasiënte vir TB-pleuritis diagnose kan verminder word, veral in MMIV.

Laastens (hoofstuk 5), het ons getoon dat in brongiale vloeistof (BF) die diagnostiese akkuraatheid nie betekenisvol verskil tussen brongoalveolêre spoelvloeistof (BALF) en brongiale spoelvloeistof (BWF) nie, en dus is hulle nie gestratifiseer in stroomaf-ontledings vir TB-diagnose nie. Ons het ook getoon dat Ultra op gekonsentreerde BF hoër sensitiwiteit en

laer spesifisiteit gehad het in vergelyking met Xpert (MIV en alternatiewe verwysingstandaarde het nie Ultra se sensitiwiteit en spesifisiteit aansienlik verander nie). Verder, 4 uit 5 Ultra "vals-positiewe" het empiriese TB behandeling begin, wat daarop dui dat Ultra op BF TB-gevalle kan opspoor wat deur kultuur gemis word. Ons het ook gewys dat programmatiese Ultra-toetsing op BF by optimalisering baat sal vind, aangesien studie-Ultra's meer TB-gevalle opgespoor het. Boonop het ons getoon dat uIFN-γ nie op BF vir TB-diagnose gebruik moet word nie weens die swak sensitiwiteit daarvan. Laastens het ons gewys dat urinêre-Ultra lae sensitiwiteit het, maar steeds TB opgespoor het wat gemis is deur toetse op die plek van siekte-vloeistof, wat die bruikbaarheid daarvan beklemtoon het.

Wat uitsette betref, het hierdie proefskrif vier eerste skrywer-manuskripte tot gevolg gehad. Een is in 'n eweknie-geëvalueerde joernaal gepubliseer (hoofstuk 2) en die ander (hoofstukke 3, 4 en 5) is gereed vir indiening. Daarbenewens word drie bykomende publikasies (waarvan een mede-eerste skrywer was) kortliks in hoofstuk 8 bespreek en kan in die bylaes gevind word. Sommige van hierdie navorsing is deur die kandidaat by 'n internasionale en nasionale eweknie-geëvalueerde konferensie aangebied.

Samevattend toon hierdie werk Ultra se hoë sensitiwiteit en matige sensitiwiteit op FNABs, perikardiale vloeistof, pleurale vloeistof en BF in pasiënte met vermoedelike TBL, TB perikarditis en TB pleuritis en PTB. Ons kan dus 'n positiewe Ultra aanbeveel met die insluiting van spoorresultate vir TB-diagnose in hierdie nie-sputummonsters.

Dissertation format

This dissertation is in the conventional format as accepted by Stellenbosch University. It includes an introduction chapter (Chapter 1), followed by 4 research chapters (Chapters 2-5), each with their own introduction, methods, results, and discussion, the first of which (Chapter 2) was published in a peer reviewed journal and co-first authored by the candidate. This is followed by a discussion highlighting the scientific contributions of the study (Chapter 6), conclusion and future work (Chapter 7), and additional academic outputs (Chapter 8). Supplementary material to research chapters and additional academic outputs are included as appendices.

Table of contents

List of abbreviations

ADA	Adenosine deaminase
ART	Antiretroviral therapy
AUROC	Area under the receiver operating characteristic
BALF	Bronchoalveolar lavage fluid
BF	Bronchial fluid (bronchoalveolar lavage fluid and bronchial wash fluid)
BWF	Bronchial wash fluid
CI	Confidence interval
Conc.	Concentrated
CRS	Composite reference standard
CSF	Cerebrospinal fluid
C_T	Cycle threshold
C_{Tmin}	Minimum cycle threshold
Culture	MGIT960 liquid culture
CXR	Chest X-ray
eMRS	Extended microbiological reference standard
EPTB	Extrapulmonary TB
FN	False-negative
FNAB	Fine needle aspiration biopsy
FP	False-positive

IFN-γ	interferon-γ
IQR	Interquartile range
KBH	Karl Bremmer Hospital
KDH	Khayelitsha District Hospital
LAM	Lipoarabinomannan
LDH	Lactate dehydrogenase
M.tb	*Mycobacterium tuberculosis*
MRS	Microbiological reference standard
MTBC	*Mycobacterium tuberculosis* complex
MTBDR*plus*	GenoType MTBDR*plus* VER 2.0
NALC-NaOH	N-Acetyl-L-Cysteine - Sodium Hydroxide
NHLS	National Health Laboratory Service
NPV	Negative predictive value
PF	Pericardial fluid (chapter 3)/ Pleural fluid (chapter 4)
PLHIV	People living with HIV
PPV	Positive predictive value
PTB	Pulmonary tuberculosis
RIF	Rifampicin
ROC	Receiver operating characteristic
SA	South Africa
SPC	Sample processing control

SR	Sample reagent
TB	Tuberculosis
TBL	Tuberculosis lymphadenitis
TB-LAM	Determine TB-LAM
TBP	TB pericarditis (chapter 3)/ TB pleuristis (chapter 5)
TGH	Tygerberg General Hospital
TN	True-negative
TP	True-positive
TTP	Time-to-positivity
uIFN-γ	unstimulated interferon-γ
Ultra	Xpert MTB/RIF Ultra
Unconc.	Unconcentrated
WHO	World Health Organization
Xpert	Xpert MTB/RIF

List of figures

Specimen concentration improves specificity in some non-sputum specimens but increases non-actionable results. Ultra has higher sensitivity compared to Xpert and uIFN-γ. Urine and TB-LAM have low diagnostic yields in non-sputum cohorts but detects patients missed by Ultra on site-of-disease fluid.

List of tables

Chapter 1

Introduction

(Literature review)

1.1 Tuberculosis and extrapulmonary tuberculosis

Tuberculosis (TB), caused by the bacterium *Mycobacterium tuberculosis* (*M.tb*), remains a leading cause of morbidity and ranks above HIV/AIDS as a leading cause of mortality[4]. The World Health Organization (WHO) estimates that in 2021, 10.6 million people fell ill from TB[4], and South Africa remains in the 30 highest TB burden countries (**Figure 1A**). TB typically infects the lungs called pulmonary TB (PTB) but can also infect other sites of the body called extrapulmonary TB (EPTB)[4]. EPTB represented 16% (~1.5 million) of all incident TB cases in 2019 (**Figure 1B**)[5]. The Covid-19 pandemic resulted in a substantial reduction in detection and reporting of TB cases in 2020 and 2021, but Africa was modestly impacted in 2020, and case detection further improved in 2021[4]. Incident TB and EPTB cases for 2022 are thus largely expected to be similar.

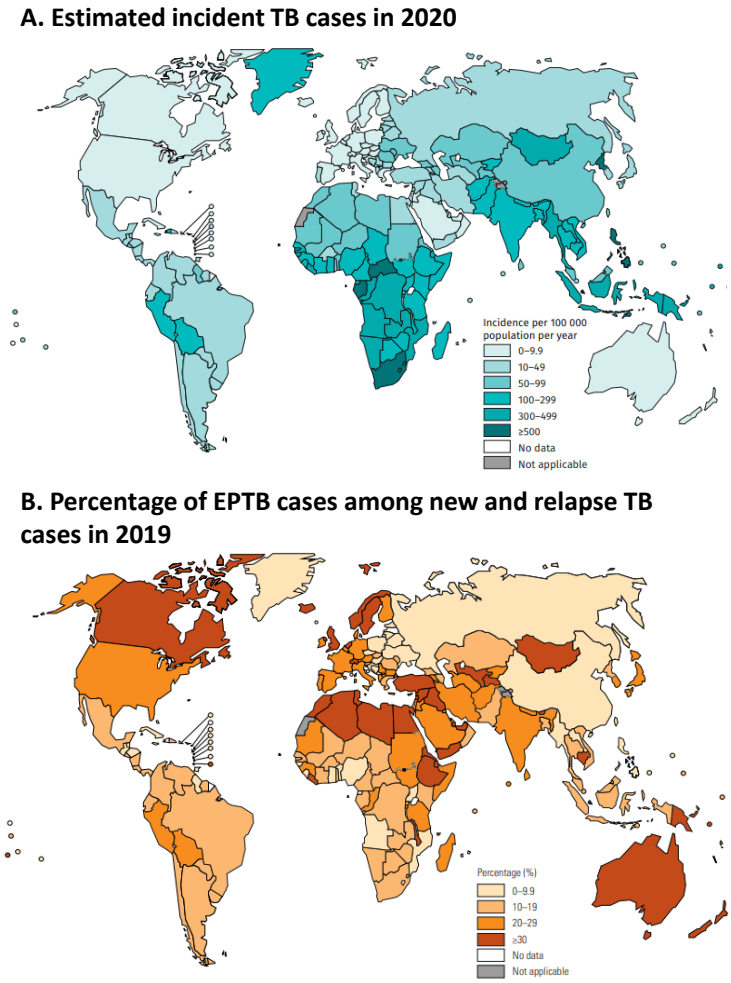


Figure 1: (A) The estimated TB incidence rates in 2020. South Africa has a high TB incidence compared to the rest of the world. **(B)** The percentage of extrapulmonary cases among new and relapse TB cases in 2019. There are a smaller number of extrapulmonary TB cases in South Africa compared to other countries, however, this number is relative to the high TB incidence in South Africa as seen in Figure 1A, estimated to be similar in 2022. Source: WHO Global Tuberculosis Report 2020 and 2022[4,5].

1.2 TB- and EPTB-HIV coinfection

Among incident TB cases in 2021, 6.7% (~710 200) were people living with HIV, with more than 50% of new TB cases in Southern Africa having an HIV co-infection[4]. People living with HIV are immunocompromised and are thus more prone to opportunistic infections such as a *M.tb* infection[6]. HIV is the strongest known risk factor for developing active TB disease from primary TB and latent TB infections in developing countries[6]. The high TB and HIV prevalence in South Africa (SA) is therefore worrisome.

The chances of developing active TB disease including EPTB is increased when HIV-coinfected[7]. This was observed in a study in Cape Town, SA which showed that when people were HIV-positive and had a CD4 cell count of <50 cells/µl, they were three-fold more likely to have EPTB compared to PTB compared to those that were HIV-negative[1]. This increased frequency of EPTB in HIV-positive patients was observed in people with varying CD4 counts (**Figure 2**)[1].

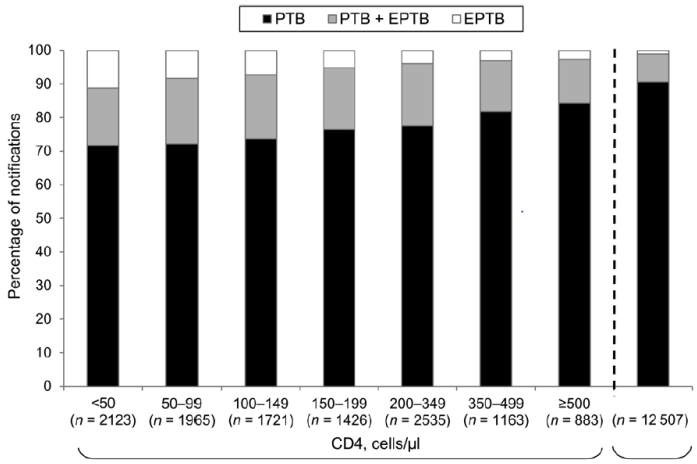


Figure 2: The increased frequency of EPTB in HIV-positive patients compared to HIV-negative patients across CD4 cell counts in a study in Cape Town, SA. Abbreviations: EPTB, Extrapulmonary TB; PTB, Pulmonary TB; SA, South Africa. Source: Gupta, *et al.*, 2013[1].

1.3 Non-sputum site-of-disease

Any site in the body can be infected with TB[8]. In contrast to symptoms typically seen for PTB (including coughing, weight loss and night sweats), EPTB symptoms can include systemic symptoms but tend to be non-specific and often are associated with other inflammatory disorders, making it challenging to recognize and diagnose[9]. Specific types of EPTB such as TB lymphadenitis (TBL), TB pericarditis and TB pleuritis in addition to PTB will be further described as they pertain to the study.

1.3.1 TB Lymphadenitis (TBL)

TBL is TB of the lymph nodes, causing swollen lymph nodes and is diagnosed by sampling and testing fine needle aspirate biopsies (FNABs) which are aspirated from the affected lymph nodes[10]. TBL is the most common manifestation of EPTB and accounts for approximately 35% (or ~525 000) of all EPTB cases[11].

1.3.2 TB pericarditis

TB pericarditis is TB of the pericardial space between the pericardium and the heart, results in pericardial fluid excess accumulation or a pericardial effusion, which is collected by a pericardiocentesis. TB pericarditis accounts for approximately 1% (or ~15 000) of EPTB cases- but is one of the deadliest forms of EPTB, and has a high fatality rate up to 40% in patients with an HIV-infection[12,13].

1.3.3 TB pleuritis

TB pleuritis is TB of the inner and outer lining of the lung or pleural space which results in the build-up of pleural fluid[14] and is collected by a thoracentesis. TB pleuritis is the second most frequent manifestation of EPTB[15] and accounts for approximately 30% (or ~450 000) of EPTB cases in TB-endemic settings[16].

1.3.4 Bronchial fluid

Bronchial fluid (BF) is bronchoalveolar lavage fluid (BALF) and bronchial wash fluid (BWF) that are pulmonary non-sputum fluids collected by a bronchoscopy by introducing saline into the lungs and then re-collecting that fluid (now containing lung cells and microbes) for analysis. Patients who require bronchoscopies often cannot produce sputum or are sputum smear-negative but are clinically suspected of having TB, and thus require this procedure[17].

Collectively, all the above non-sputum specimens tend to be paucibacillary and often require invasive sampling from clinicians with expertise and significant infrastructure[3]. These factors combined with an HIV-co-infection affects the diagnostic accuracy of microbiological and molecular tests. Thus, the diagnosis of these specimens remains a challenge[9].

1.4 Current diagnostic challenges for non-sputum fluids

1.4.1 Culture

TB culture, particularly liquid culture, or Mycobacteria Growth Indicator Tube (MGIT960) remains the gold standard for diagnosing TB, however, it takes 4-6 weeks for a result. Culture has been shown to have a sensitivity between 60-70% for TBL, 53% for pericardial fluid and 30-50% for pleural fluid when compared to composite clinical reference standards (typically including microbiological and cytological readouts from beyond the site of disease and/or clinical information; see cohort specific chapters for precise definitions), highlighting its suboptimal sensitivity in non-sputum specimens[3,18]. The use of nucleic acid amplification tests such as Xpert MTB/RIF were therefore developed to improve TB detection in these specimens[3].

1.4.2 Xpert MTB/RIF (Xpert)

Xpert MTB/RIF (Xpert) (Cepheid, USA) is an automated real-time PCR system, which rapidly detects *M.tb* complex (MTBC) DNA and rifampicin resistance, and was developed to improve *M.tb* detection[19,20]. Xpert has been shown to perform poorly in smear-negative and EPTB specimens (with sensitivity as low as 48%)[21], nevertheless, it is an advance over smear microscopy which although rapid and inexpensive, has extremely low sensitivity in pulmonary and EPTB specimens[22,23]. Moreover, a systematic review showed large heterogeneity in the sensitivity of Xpert in different non-sputum specimens such as pleural fluid when compared to culture (with sensitivity ranging from 0-100%) (**Figure 3A**), or when compared to a composite reference including culture and smear (with sensitivity ranging from 13-50%) (**Figure 3B**)[3]. This heterogeneity in sensitivity was similarly observed in other non-sputum specimens such as FNABs (50-100%)[3] and cerebrospinal fluid (CSF) (0-100%)[3] as well as a more recent systematic review[24]. A more sensitive and consistent diagnostic test was therefore needed to improve the detection of *M.tb* in non-sputum specimens.

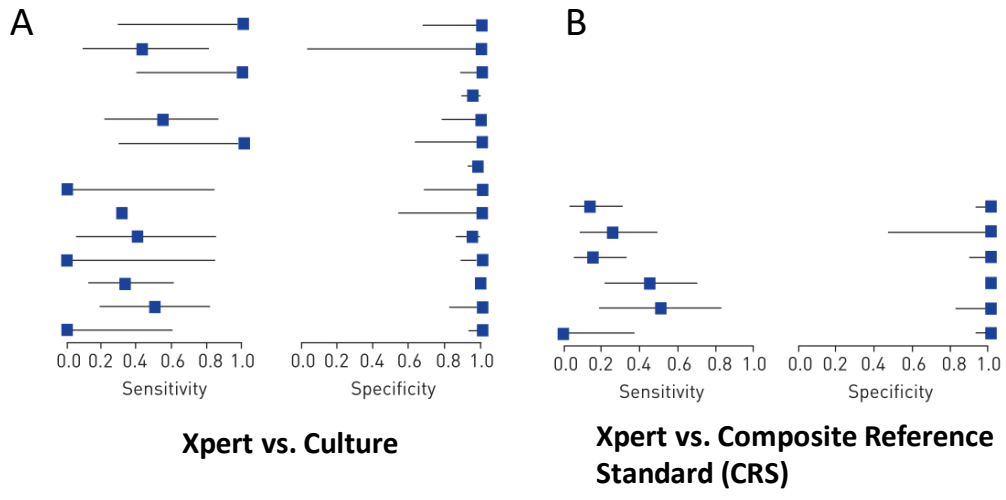


Figure 3: **(A)** Forest plot showing the heterogeneity in sensitivity of Xpert when pleural fluid was tested in various studies and compared to culture. (**B**) Forest plot showing that when Xpert was compared to a composite reference standard (which may have included smear, histology and presenting symptoms and response to TB treatment), heterogeneity in sensitivity was still observed. This range in sensitivity was similarly observed in other non-sputum specimens. Source: Denkinger *et al.*, 2014[3].

1.4.3 Xpert MTB/RIF Ultra

Xpert MTB/RIF Ultra (Ultra) is the successor to Xpert offering improved sensitivity for MTBC DNA. This is partly enabled by a larger reaction chamber that doubles the real-time PCR reaction volume compared to Xpert, and the targeting of two *M.tb*-specific multi-copy genes (*IS*6110, *IS*1081) in addition to *rpoB*, whereas Xpert just uses a single copy *rpoB* gene target for MTBC detection and rifampicin resistance determination[2]. The limit of TB detection for Xpert was shown to be 112.6 colony forming units/ml (CFU/ml) (**Figure 4A)** compared to the 15.6 CFU/ml for Ultra (**Figure 4B**) when sputum spiked with *M.tb* DNA was tested[2]. This suggests that Ultra could result in greater TB detection in non-sputum specimens with or without a HIV co-infection. At the start of the study, Ultra had not been evaluated in any fluid type, particularly in HIV-endemic settings, highlighting the need for this data.

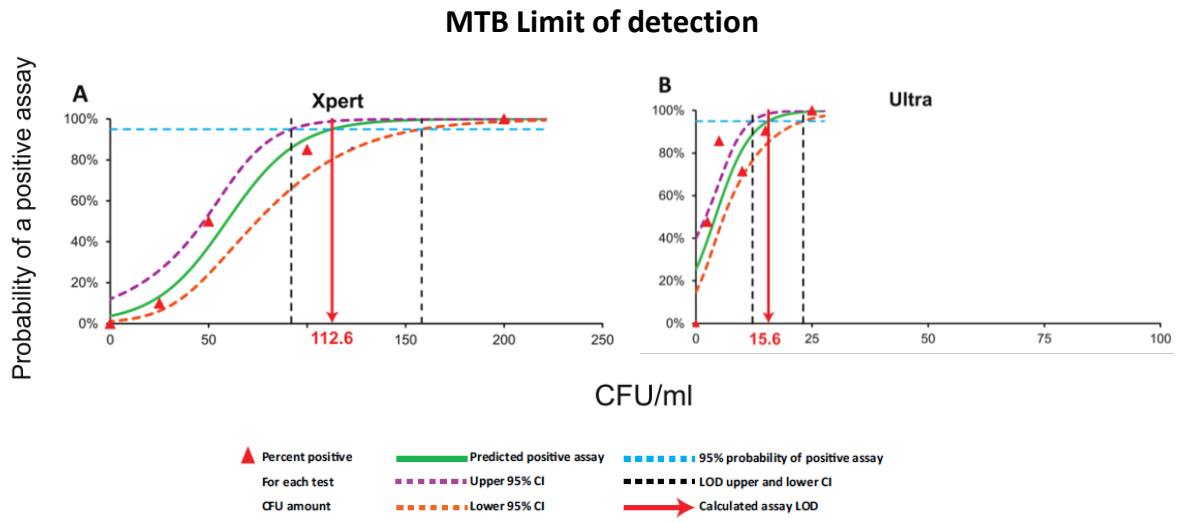


Figure 4: (**A**) The limit of detection of Xpert in *M.tb* spiked sputum is 112.6 CFU/ml. (**B**) The limit of detection of Ultra is 15.6 CFU/ml and improved compared to Xpert. Ultra in non-sputum specimens is likewise expected to have improved limit of detection, but this has not been verified. Source: Chakravorty *et al.*, 2017[2].

1.4.4 Unstimulated interferon gamma (uIFN-γ) in site-of-disease fluid

Interferon gamma (IFN-γ) is an immunological cytokine produced in the body in response to infection. Elevated levels of IFN-γ indicates the body's response to a non-specific bacterial or viral infection[25]. Interferon gamma release assays (IGRAs) are *in vitro* blood tests that measures the release of IFN-γ following stimulation of MTBC antigens[26], however the WHO strongly discourages the use of IFN-γ release assays for the diagnosis of any form of active extrapulmonary TB[27].

Unlike a stimulated IFN-γ release assay, unstimulated interferon gamma (uIFN-γ), measured using an enzyme-linked immunosorbent assay (ELISA), has shown high sensitivity and specificity in distinguishing non-TB cases from TB cases in non-sputum specimens such as pleural fluid[28], even while not specifically measuring an *M.tb* infection. This was observed in intermediate and low TB burden settings, and later a high TB prevalence setting such as South Africa[29]. This suggests that uIFN-γ levels holds promise for differentiating TB cases in non-sputum specimens.

1.4.5 Determine TB-LAM (TB-LAM) and Ultra on urine

The use of an easily accessible fluid like urine could mitigate the need for invasive sampling procedures and specialised equipment often needed for the collection of non-sputum specimens. This is particularly essential for EPTB and PTB diagnosis (in sputum-scarce patients). Determine TB-LAM (TB-LAM; Abbott, USA) is a lateral flow assay that detects lipoarabinomannan (LAM), a mycobacterial cell wall component that accumulates in urine[30] and is endorsed by the WHO for people living with HIV[5]. A systematic review and meta-analyses of the diagnostic yield of TB-LAM in patients with PTB showed that it is effective for TB diagnosis in PLHIV, particularly in people with low CD4 counts or inpatients[31]. Like TB-LAM, Ultra on urine could improve TB diagnosis as seen in patients with renal TB or TB

meningitis[32,33] or PTB[34]. Moreover, urine concentration by centrifugation significantly increased the sensitivity of Xpert over unconcentrated urine in sputum scarce patients that were HIV-postive[35]. These data suggests that TB-LAM on urine and Ultra on concentrated urine could be useful for the diagnosis of PTB and EPTB when not detected by tests on site-of-disease fluid.

2. Key knowledge gaps

This work addressed knowledge gaps associated with the performance of tests that, at the advent of the study, were new and novel, with a focus on test performance on non-sputum specimens, especially those from people with presumptive EPTB. It accomplished this by leveraging a large referral network established at our teaching (and surrounding) hospitals, enabling it to mostly recruit large cohorts of patients, evaluating these tests on site-of-disease fluids (often before those tests were, if at all, offered by the programme as well as urine). The specific gaps and associated research questions include:

2.1. What is the diagnostic accuracy of Ultra on fine needle aspirate biopsies (FNABs) from lymph nodes, pericardial fluid, pleural fluid, and bronchial fluid [bronchoalveolar lavages and bronchial washes] in different adults with presumptive EPTB or PTB?

There have been few studies investigating the performance of Ultra in non-sputum specimens published to date, however, the sensitivity of Ultra is promising[2]. Ultra has been shown to have high sensitivity for the diagnosis of TBL when FNABs in SA were compared to a composite reference standard (CRS)[36], when Ultra was evaluated in pleural fluid and compared to culture in China[37] and South Africa[38], and, when Ultra was compared to Xpert in BF. Ultra also showed promise when a combination of smear-negative non-sputum specimens (including fine needle biopsies, pericardial fluid and pleural fluid) were tested[19,39,40]. It is important to note that at the start of the study, no studies evaluating Ultra on non-sputum specimens existed. Moreover, no study at present has evaluated the same cohort of non-sputum specimens against different reference standards. This suggests that Ultra could be a highly sensitive diagnostic tool for testing non-sputum specimens, however, more data is needed to confirm these findings, particularly in an HIV-endemic setting such as South Africa.

2.2. How do different sample processing methods such as concentration of fluid affect the diagnostic accuracy and non-actionable rate of Ultra?

It remains uncertain how well various sample processing methods such as sample concentration of non-sputum specimens affects the diagnostic accuracy and moreover, non-actionable rate of Ultra. Sample concentration by centrifugation has been previously shown to improve diagnostic accuracy of Xpert when non-sputum specimens like urine, blood and CSF were tested[35,41,42]. To date, the effect of sample concentration to improve Ultra sensitivity has only been described in CSF[42], thus more data on the effect of concentration of other non-sputum specimens such as pericardial, pleural and bronchial fluid is needed. Non-sputum specimens are often concentrated by centrifugation in routine laboratories to increase the concentration of target analytes for improved test detection, but it is unknown if this aids the diagnostic accuracy of Ultra or hinders by resulting in more Ultra non-actionable results (since other components of the specimen are also concentrated which can inhibit the test). Knowing this could inform guidelines that can easily be implemented for optimal processing of non-sputum specimens.

2.3 How does the diagnostic accuracy of Ultra compare to that of Xpert and unstimulated interferon gamma ELISA (uIFN-γ) on different non-sputum specimens?

Although Ultra has been endorsed by the WHO for use in PTB and rolled out globally, limited data on head-to-head diagnostic accuracy evaluations of Ultra compared to Xpert exists for non-sputum specimens[20,37,43], and no data exists comparing the diagnostic accuracy of Ultra to uIFN-γ on any non-sputum specimen. This is important as uIFN-γ has been shown to have high diagnostic accuracy in pleural fluid[29,38] and pericardial fluid[44].

2.4 How does the diagnostic yield of Ultra on site-of-disease fluid compare to that of TB-LAM and Ultra on urine?

The diagnostic yield of TB-LAM has been evaluated in a systematic review and meta-analysis in people living with HIV (PLHIV) in patients with PTB[31]. Moreover, both TB-LAM and Ultra on urine were compared in patients with PTB[34], in a patient with renal TB[32] and in patients with suspected TB meningitis[33]. At present, no study has evaluated the diagnostic accuracy of both TB-LAM and Ultra on urine in patients with presumptive TBL, TB pericarditis and TB pleuritis in an HIV-endemic setting. Furthermore, tests on urine have not been directly compared to Ultra on site-of-disease fluid for EPTB or PTB diagnosis. With this knowledge, patients could potentially be diagnosed without the need for invasive sampling.

3. Study rationale and concluding remarks

In summary, Ultra is endorsed for TB diagnosis on sputum but, at the study's advent, more data was needed regarding usefulness in paucibacillary non-sputum specimens, particularly in HIV-endemic settings, and this need persists to this day. Additionally, knowledge of the optimal sample processing method for Ultra could decrease non-actionable results and inform local clinical practice and policy guidelines recommendations. The need for sensitive tests that uses non-sputum specimens has also been emphasised by the WHO, highlighting the need for data on an easily accessible fluid such as urine in cohorts not previously investigated. Moreover, data for how Ultra directly compares to other tests on site-of-disease non-sputum specimens and tests on urine could affect global diagnostic guidelines.

4. Summary of knowledge gaps and aims (Figure 5)

Our overarching aim is to evaluate the diagnostic accuracy of Ultra on fine needle aspirate biopsies (FNABs) on lymph nodes, pericardial fluid, pleural fluid and bronchial fluid in adults with presumptive EPTB or PTB respectively, undergoing routine clinical investigation.

Aim 1: To evaluate the diagnostic accuracy of Ultra in different non-sputum specimens (FNABs, pericardial fluid, pleural fluid, and bronchial fluid) using an unconcentrated and, when available, concentrated specimen for each fluid type using MGIT960 liquid culture as a microbiological reference standard (MRS).

Sub-aim 1.1: To assess the impact of sample concentration on the diagnostic accuracy and non-actionable rate of Ultra for each fluid type.

Sub-aim 1.2: To evaluate the diagnostic accuracy of Ultra for each fluid type using an extended microbiological reference standard (eMRS) and composite reference standard (CRS).

Sub-aim 1.3: To evaluate whether Ultra minimum cycle threshold correlates with bacterial load [i.e., Ultra C_{Tmin} vs. culture time to positivity (TTP)].

Sub-aim 1.4: To compare Ultra diagnostic accuracy performance by patient sub-groups for each fluid (i.e., HIV and previous TB status).

Sub-aim 1.5: To compare the characteristics of true-positive and false-positive Ultra results for each fluid (i.e. previous TB status, HIV status, CD4 count).

Aim 2: To compare the diagnostic accuracy of Ultra with other tests such as Xpert and uIFN-γ on site-of-disease fluid in each non-sputum cohort (FNABs, pericardial fluid, pleural fluid and bronchial fluid) compared to the MRS.

Sub-aim 2.1: To compare the diagnostic accuracy of Ultra with Xpert on site-of-disease fluid in each fluid type compared to alternate standards (eMRS and CRS).

Sub-aim 2.2: To determine the optimal rule-in rule-out cut-points for uFNγ for each fluid type using a receiver operating characteristic (ROC)-curve analysis.

Sub-aim 2.3: To compare the diagnostic accuracy of Ultra with uIFN-γ on site-of-disease fluid in each fluid type compared to alternate standards (eMRS and CRS).

Aim 3: To determine the diagnostic yield of Ultra and TB-LAM on urine in different non-sputum cohorts (FNABs, pericardial fluid, pleural fluid, and bronchial fluid).

Sub-aim 3.1: To compare the diagnostic accuracy of Ultra on site-of-disease fluid with tests on urine (i.e., Ultra and TB-LAM) in each fluid type compared to the MRS.

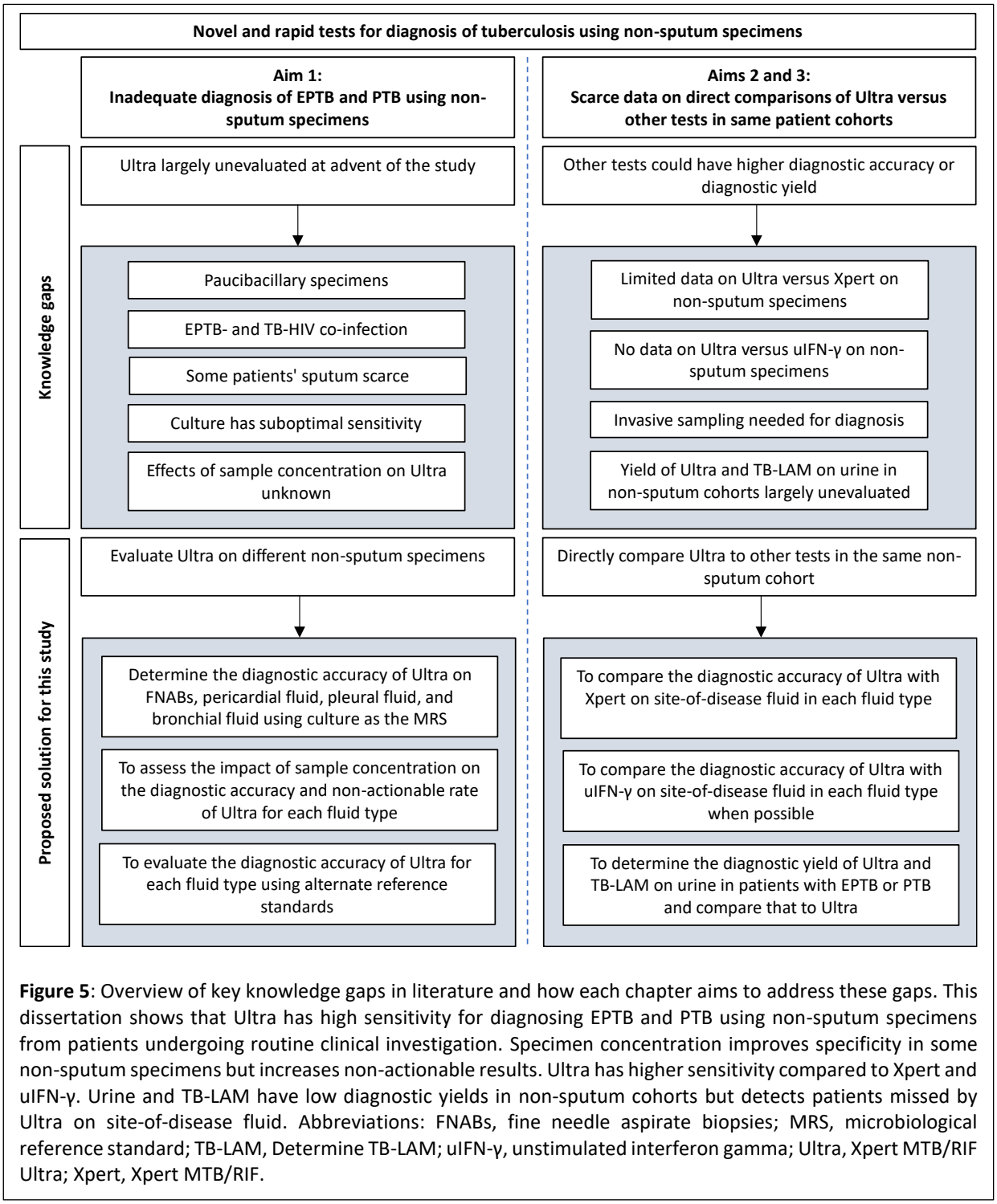


Figure 5: Overview of key knowledge gaps in literature and how each chapter aims to address these gaps. This dissertation shows that Ultra has high sensitivity for diagnosing EPTB and PTB using non-sputum specimens from patients undergoing routine clinical investigation. Specimen concentration improves specificity in some non-sputum specimens but increases non-actionable results. Ultra has higher sensitivity compared to Xpert and uIFN-γ. Urine and TB-LAM have low diagnostic yields in non-sputum cohorts but detects patients missed by Ultra on site-of-disease fluid. Abbreviations: FNABs, fine needle aspirate biopsies; MRS, microbiological reference standard; TB-LAM, Determine TB-LAM; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

5. Originality of the study and impact

There are studies that have evaluated the diagnostic accuracy of Ultra on non-sputum specimens including FNABs[36], pleural fluid[38,45] and bronchoalveolar lavage fluid[43,46] since the start of the study. These studies however did not compare the effect of sample processing such as concentration and the use of different reference standards on the diagnostic accuracy of Ultra. Furthermore, while there have been studies evaluating the diagnostic accuracy of Ultra on a limited combination of EPTB fluids that included pericardial fluid, a large diagnostic accuracy study evaluation of Ultra on pericardial fluid only, particularly in an HIV-endemic setting, has not been done before.

Additionally, many of the abovementioned studies did not directly compare the diagnostic accuracy of Ultra with its predecessor, Xpert or with other tests like uIFN-γ on site-of-disease fluid. The diagnostic yield of TB-LAM and Ultra on urine has not been evaluated in patients with presumptive TBL, TB pericarditis and TB pleuritis before. Moreover, Ultra's TB detection on site-of-disease fluid has not been compared to TB-LAM and Ultra on non-site-disease-fluid such as urine before. Our study will therefore create valuable, much needed data on diagnostic accuracy of Ultra on different non-sputum specimens. These data will correspondingly inform local clinical practice such as potential improvements of existing SA National Health Laboratory Service (NHLS) non-sputum laboratory processing and testing systems, or confirmation of extant systems. This study will therefore contribute towards national policy guideline recommendations, which could affect global diagnostic guidelines.

Chapter 2

Xpert MTB/RIF Ultra Is Highly Sensitive for the Diagnosis of Tuberculosis Lymphadenitis in a High-HIV Setting

Supplementary material is attached as Appendix I

Publication status: published

Minnies, S*, Reeve, B.W.*, Rockman, L., Nyawo, G., Naidoo, C.C., Kitchin, N., Rautenbach, C., Wright, C.A., Whitelaw, A., Schubert, P., Warren, R.M., Theron, G., 2021. Xpert MTB/RIF Ultra is highly sensitive for the diagnosis of tuberculosis lymphadenitis in a High-HIV setting. *Journal of Clinical Microbiology*, 59(12), pp.10-1128.

DOI: https://doi.org/10.1128/jcm.01316-21

****Contributed equally***

Take home message:

Ultra on FNABs is highly sensitive and detects more TBL than Xpert and would result in more people placed on TB treatment, and this is driven by the added benefit of trace results. Urine-Ultra could reduce the number of patients needing invasive sampling and associated delays.

Candidate's role:

Processing of Ultras on FNABs, data analysis and preparation of manuscript.

Citations: 16

Abstract

Background: Tuberculosis lymphadenitis (TBL) is the most common extrapulmonary TB (EPTB) manifestation. Xpert MTB/RIF Ultra (Ultra) is a World Health Organization-endorsed diagnostic test, but performance data for TBL, including on non-invasive specimens, are limited.

Methods: Fine needle aspiration biopsies (FNABs) from outpatients (≥18 years) with presumptive TBL (n=135) underwent: 1) routine Xpert (later Ultra once programmatically available), 2) a MGIT960 culture (if Xpert- or Ultra-negative, or rifampicin-resistant), and 3) study Ultra. Concentrated paired urine underwent Ultra. Primary analyses used a microbiological reference standard (MRS).

Results: In a head-to-head comparison (n=92) of FNAB study Ultra and Xpert, Ultra had increased sensitivity [91% (95% confidence interval 79, 98) vs. 72% (57, 84); p=0.016] and decreased specificity [76% (61, 87) vs. 93% (82, 99); p=0.020], and detected patients not on treatment. HIV nor alternative reference standards affected sensitivity and specificity. In patients with both routine and study Ultras, the latter detected more cases [+20% (0, 42); p=0.034] and, further indicative of potential laboratory-based room-for-improvement, false-negative study Ultras had more PCR inhibition than true-positives. Study Ultra "false-positives" had less mycobacterial DNA than "true-positives" [trace-positive proportions 59% (13/22) vs. 12% (5/51); p<0.001]. Exclusion or recategorization of "traces" removed potential benefits offered over Xpert. Urine-Ultra had low sensitivity [18% (7, 35)].

Conclusions: Ultra on FNABs is highly sensitive and detects more TBL than Xpert. Patients with FNAB Ultra-positive "trace" results, most of whom will be culture-negative, may require additional clinical investigation. Urine-Ultra could reduce the number of patients needing invasive sampling.

247/250

Background

Tuberculosis (TB) is a leading cause of morbidity and mortality globally. In 2019, extrapulmonary TB (EPTB) represented 16% of new TB cases reported[5] and, in HIV-positive populations, can account up to 50% of all TB[8]. TB lymphadenitis (TBL) accounts for 35% of all EPTB[47,48]. South Africa, with its' high TB and HIV burden[5], is particularly affected by EPTB and TBL.

TBL is typically diagnosed by examining fine needle aspiration biopsies (FNABs) from affected lymph nodes. This requires specialised sampling and facilities, and tests have suboptimal sensitivity[21]. One widely-used test is Xpert MTB/RIF (Xpert; Cepheid, USA); a semi-automated real-time PCR that rapidly detects *Mycobacterium tuberculosis* complex (MTBC) DNA and rifampicin resistance[2,49]. A systematic review and meta-analysis showed heterogeneity in the sensitivity of FNAB Xpert vs. microbiological [83% (95% confidence interval: 71, 91) and composite reference standards [81% (72, 88)][3]. Specificities were 94% (88, 97) and 99% (95, 100), respectively[50]. Most EPTB diagnostic algorithms recommend culture after a negative Xpert[51], however, this creates delay. Better TBL tests are needed.

One potential test is Xpert MTB/RIF Ultra (Ultra), which offers improved sensitivity over Xpert for pulmonary TB, partly enabled by, in addition to *rpoB*, amplification of multi-copy insertion elements (IS*6110,* IS*1081*)[52]. Data on Ultra for TBL are emerging: one retrospective evaluation tested ten Xpert-negative, culture-positive FNABs and found half to be Ultra-positive[19], another retrospective evaluation (n=25) reported sensitivity and specificity of 94% (71-77) and 100% (63-72), respectively[39]; and a prospective evaluation (n=73) reported a sensitivity and specificity of 78% (40-97) and 78% (66-87), respectively[36]. No studies included head-to-head Xpert and Ultra data. Additionally, since Ultra's advent, algorithms for TBL diagnosis remain essentially unchanged from the Xpert era – culture is still recommended in

Ultra-negative patients. Whether this is needed or, conversely, if culture is needed to confirm positive Ultra results due to specificity concerns associated with the new trace semi-quantitation category[52,53], requires investigation.

Lastly, FNABs are rarely collected in primary care; patients are referred to district or tertiary facilities, resulting in care cascade gaps[54]. If an Ultra has high sensitivity and specificity on an easily accessible fluid like urine, the need for invasive sampling could be mitigated; potentially drastically reducing provider and patient economic and time costs. To our knowledge, urine-Ultra for TBL is unevaluated.

We evaluated the head-to-head diagnostic accuracy of Xpert and Ultra on FNABs, and Ultra on urine in patients with presumptive TBL in a tertiary hospital setting in an HIV-endemic in South Africa. We hypothesised Ultra would show improved sensitivity compared to Xpert.

Methods and materials

Ethics statement

The study was approved by the Stellenbosch University Human Research Ethics Committee and Tygerberg General Hospital (TGH) (both N16/04/050).

Patient recruitment

135 outpatients (≥18 years) with presumptive TBL (swollen lymph node) undergoing routine referral and investigation at a tertiary referral clinic at TGH in Cape Town, South Africa, were consecutively recruited from 25 January 2017-12 March 2019 and gave FNABs and urine. Patients who received TB treatment ≤60 days prior were excluded.

Fine needle aspirate collection

FNABs were collected by multiple needle passes using a 23-gauge needle and 10 ml syringe. While the needle was inserted, negative suction with a cutting motion was applied for aspiration. The first two passes were used for routine cytology. From each pass, two slides were prepared: the first airdried for Rapidiff staining and the second spray-fixed for Papanicolaou staining (~25 µl total volume used per pass) (**Figure 1**). The remaining syringe contents were flushed into 1.5 ml TB transport medium[55]. The third pass (5-50 µl) was collected into 700 µl 5% saline (Ysterplaat Medical Supplies, Cape Town, South Africa).

Xpert, Ultra, and culture

Routine testing: Xpert (version 1; Cepheid, USA) was done programmatically from 25 January 2017–9 April 2018 by the government programmatic laboratory [National Health Laboratory Service (NHLS)] who did Ultra (version 1) thereafter[56]. Sample reagent (2 ml; Cepheid, USA) was added to 500 µl of aspirate-containing 1.5 ml TB transport medium (4:1 ratio) and 2 ml of the mixture used for Xpert or Ultra[57,58]. Per the algorithm, if a specimen was Xpert- or Ultra-positive and rifampicin-susceptible, culture was not done. If Xpert- or Ultra-negative, or Xpert-

or Ultra-positive and rifampicin-resistant, 500 µl aspirate-containing TB transport medium was inoculated into a MGIT960 liquid culture without NALC-NaOH decontamination (**Figure 1**). If a non-actionable (not positive or negative)[53] Xpert or Ultra occurred, the remaining 500 µl TB transport medium was used to repeat the test.

Study testing: The third pass in 700 µl saline was tested with Ultra (version 3; study Ultra) using a 2:1 sample reagent ratio[57]. Study Ultra was done irrespective of whether routine Xpert or Ultra was done.

MTBC typing and drug susceptibility testing: MTBDR*plus* was done on culture-positive isolates for speciation and drug susceptibility testing.

Urine-Ultra

5-20 ml urine stored at -80 ºC were centrifuged (1811×*g*, 10 min, room temperature) and the supernatant removed until 700 µl remained, which was tested with Ultra (2:1 sample reagent volume ratio)[58].

Patient treatment and follow-up

Treatment decisions were programmatic without study involvement (no study results reported for patient management). Attempts were made to telephonically follow-up patients at least 12 weeks after recruitment at which point TB treatment initiation status were recorded and, if treatment started, treatment response was queried. Patients were lost-to-follow-up if at least two calls were unsuccessful, and messages were unreturned for each timepoint.

Definitions

Patient groups: Patients were designated definite, probable, or non-TB using different reference standards. For the microbiological reference standard (MRS), definite TB was culture-positive and/or cytology-positive on FNABs, and non-TBs culture- and cytology-negative on FNABs. Unclassifiable patients had no positive MRS test, culture contaminated or

not done, and cytology not done. **Supplementary Table 1** has the extended microbiological standard (eMRS) and composite reference standard (CRS) definitions.

Other definitions: Xpert or Ultra actionable results for TB were MTBC-detected and rifampicin-susceptible, rifampicin-resistant or rifampicin-indeterminate, or MTBC not detected. For culture, actionable results were positive or negative for MTBC. For cytology, the presence or absence of granulomatous inflammation was recorded.

Statistical analysis

We included patients in head-to-head analyses if they had actionable routine index test (Xpert or Ultra), study Ultra, and culture results (or, if culture was non-actionable, a cytology result was available). Proportion tests[59] were done using STATA version 16.0 (StataCorp, College Station Texas, USA) and GraphPad Prism version 8.0.1 (GraphPad Software, San Diego, USA). Venn diagrams were made with InteractiVenn[60]. Differences in diagnostic accuracy metrics were calculated using proportion tests or McNemar's test as appropriate. STARD guidelines were followed[61].

Results

Patient characteristics

Of 135 patients, 44% (59/135) were definite TB and 56% (75/135) non-TBs per the MRS. Characteristics are compared in **Table 1**.

FNAB index test results

76% (103/135) of patients had routine Xpert requested [6/103 (6%) not done] and 24% (32/135) routine Ultra requested [3% (1/32) not done]. Non-actionable results for routine Xpert, routine Ultra and study Ultra were 0% (0/97), 6% (2/31), and 3% (4/135), respectively. 41% (40/97) of routine Xperts were positive (remainder negative). For routine Ultra, 38% (11/29) were positive and, for study Ultra, 74/131 positive (56%; p=0.070 vs. routine Ultra) (**Figure 2**). In a head-to-head comparison of patients with actionable results from each test (study Ultra, routine Xpert, culture, cytology) 37% (22/59), 8% (5/59), 20% (12/59) and 24% (14/59) were positive by each test (**Figure 3A**; study Ultra had the highest yield). 12% (7/59) of these patients with at least one positive result were exclusively detected by study Ultra (cytology exclusively detected two). This proportion detected only by study Ultra (and hence were negative by routine Xpert and/or cytology) increased to 22% (13/59) when culture results, which are not available for rapid clinical decision making, were omitted.

Diagnostic accuracy and yield of study Ultra and routine Xpert on FNABs

Overall: When Ultra was compared head-to-head to Xpert using the MRS (n=92) (**Table 2**), Ultra had improved sensitivity [91% (95% confidence interval: 79, 98) vs. 72% (57, 85); p=0.016] and decreased specificity [76% (61, 87) vs. 93% (82, 99); p=0.020]. Ultra's positive predictive value (PPV) [79% (66, 89) vs. 92% (78, 98); p=0.114] and negative predictive value (NPV) were like Xpert's [90% (76, 97) vs. 77% (64, 87); p=0.105]. Conclusions were unchanged for non-head-to-head comparisons, eMRS or CRS (**Table 2, Supplementary results**). Compared to MTBDR*plus* on isolates, no false-negative or false-positive Ultra

rifampicin-resistance results occurred, however, numbers were small, precluding precise accuracy estimates (**Supplementary Results**).

HIV: Sensitivities and specificities did not differ in HIV-positives vs. -negatives for study Ultra or routine Xpert (**Table 2**). Within HIV-positives, Ultra had improved sensitivity [97% (82, 100) vs. 76% (56, 96); p=0.022] and similar specificity [79% (59, 92) vs. 93% (76, 99); p=0.127] to Xpert.

Trace semi-quantitation exclusion or reclassification: When study Ultra traces were excluded, sensitivity [-1% (-17, 11); p=0.836] and specificity [+7% (-9, 24); p=0.400] were unchanged. When trace results were reclassified as negative, sensitivity decreased [-13% (-25, 1), p=0.014] and specificity increased [+9% (-2, 19), p=0.046] (**Table 2**).

Ultra PCR inhibition: An analysis of sample processing control (SPC) C_T values (**Supplementary Figure 1**; higher values indicate more inhibition) showed more inhibition in study Ultra positives than -negatives [25.80 (IQR: 24.78-27.33) vs. 25.20 (24.55-26.05); p=0.024]. Furthermore, false-negatives were more inhibited than true-positives [26.10 (25.10-28.60) vs. 25.10 (24.00-25.50); p=0.001]; suggesting inhibition contributes to diminished sensitivity.

Relationship with bacterial load: Neither Study Ultra nor routine Xpert C_T correlated with bacillary load measured using culture time-to-positivity (**Supplementary Figure 2**) in FNABs.

<u>Comparison of study Ultra true-positive and false-positives</u>

False-positives had less bacterial load than true-positives [IS*6110*/IS*1081* C_T 19.00 (IQR: 16.40-21.60) vs. 24.85 (19.88-28.15); p<0.001], a greater proportion were hence “trace” [59% (13/22) vs. 12% (6/51); p<0.001] (**Table 4**). Less inhibition was also observed for the former group [SPC C_T 25.05 (24.45-25.95) vs. 26.10 (25.10-28.60); p=0.005]. More study Ultra true-positives were on treatment at follow-up than Ultra false-positives [92% (44/48) vs. 27% (6/22); p<0.001] as more true-positives were positive using a routine test than the false-

positives [98% (50/51) vs. 27% (6/22); p<0.001]. The proportions of patients with previous TB in false- vs. true-positives were similar [27% (6/22) vs. 35% (18/51); p=0.503]. The characteristics of true- and false-positives are in **Table 4** and false-positives per patient information in **Supplementary Table 3**.

Study vs. routine Ultra FNAB results

Concordance: In patients who received both study and programmatic Ultras, 55% (17/31) were study Ultra-positive and 35% (11/31) routine Ultra-positive. The former detected +20% (95% confidence interval: 0, 42) more TBL (**Table 3**).

PCR inhibition: SPC C_T analysis showed no difference between study and routine Ultra [25.10 (IQR: 24.35-25.85) vs. 25.50 (24.20-26.50); p=0.081] (**Supplementary Figure 1A**).

Urine-Ultra yield, sensitivity and specificity, and non-actionable results

Urine-Ultra had low sensitivity [18% (7, 35)] and high specificity [98% (88, 100)] (head-to-head comparisons with FNAB study Ultra in **Supplementary Table 4**). Of concentrated urines tested with (n=84), 8% (7/84) were non-actionable and 100% (7/7) of these resolved to actionable when unconcentrated urine was tested (one unconcentrated urine was now Ultra-positive). None of the 18 HIV-negative patients had any positive urine-Ultra. 12% (7/57) HIV-positives were urine-Ultra-positive (six of seven detected by both positive MRS and study Ultra FNAB result; **Figure 3C**). In other words, when urine-Ultra was attempted amongst HIV-positives, 11% (7/64, 3 of which were trace) were positive, meaning that universal concentrated urine-Ultra testing in HIV-positives with presumptive TBL could reduce the number of FNABs required for TB diagnosis as few are non-actionable.

Patient treatment status at follow-up

96% (130/135) of patients were followed-up [median (IQR: 37 (16-65) weeks since recruitment] and 52% (68/130) of those had initiated treatment. Of these, 74% (50/68) had been

classified as definite TB and 26% (18/68) as non-TB per the MRS. Of the definite TBs, 88% (44/50) were study Ultra-positive whereas, for the non-TBs, 33% (6/18) were study Ultra-positive. Of the remaining study Ultra-positives followed-up, 29% (20/70) were not placed on treatment [in 65% (13/20) of these, study Ultra was the only positive test], indicating potential missed opportunities for treatment initiation. Regarding the clinical status in patients who started treatment, 94% (64/68) reported treatment completion and, of these, 94% (60/64) reported feeling clinically well. 3% (4/130) patients were documented to have died (one of the four had a positive test result that was exclusively study Ultra positive; none of these four were placed treatment).

Discussion

Our key findings are: 1) study Ultra on FNABs had, compared to Xpert, improved sensitivity and decreased specificity, and outperformed routine Ultra (tests unaffected by HIV and alternative reference standards); 2) approximately 3 in 10 study Ultra-positives had not been placed on treatment, indicating opportunities to improve TBL treatment with Ultra; 3) excluding study Ultra trace results improved specificity (more so than reclassifying to negative) without large sensitivity costs relative to treating Ultra trace results as positive; 4) Urine-Ultra had low sensitivity but could reduce the proportion of presumptive TBL patients who require a FNAB in our setting, and 5) Ultra false-negative results are associated with PCR inhibition. These data show high sensitivity of Ultra on FNABs for TBL with the inclusion of trace-positive results (without which sensitivity benefits over Xpert are not seen).

Ultra on FNABs had increased sensitivity than Xpert, suggesting Ultra is rapid initial test for TBL. Ultra did still not detect, however, approximately 1 in 10 TBL cases; indicating a sustained need for more sensitive tests (especially those that use non-invasive specimens) and a continued role for reflex tests for downstream testing of Ultra-negative FNABs. Importantly, like was done previously for Xpert[62], we showed one likely cause of Ultra false-negativity is increased PCR inhibition, suggesting that optimised specimen processing workflows (such as centrifugation and removal or dilution of inhibitor containing supernatant)[42] to better remove interfering agents are still needed to boost sensitivity.

Notably, Ultra had suboptimal specificity (two in ten MRS-negative people were study Ultra-positive). One reason may be that culture and cytology have limitations as reference standards for EPTB[3]. Notably, this finding mirrors prior work on TBL that used tissue in addition to fluid biopsies for Ultra, where a specificity of 78% vs. culture was observed[36]. However, when compared to an eMRS including microbiological tests such as FNAB culture as well as culture and Ultra on non-site-of-disease fluids, FNAB Ultra specificity was 100% in that study. In

contrast, we applied microbiological tests only to FNABs and did not exhaustively sample anatomical sites[63], which might underestimate specificity.

Ultra false-positivity was more frequent in patients with less mycobacterial DNA and, in contrast to pulmonary TB, FNAB Ultra false-positivity was not associated with prior TB[53]. The true nature of these Ultra "false-positives" in EPTB requires clarification and is an important topic for future research (in our setting, most "false-positive" patients with presumptive pulmonary TB remain well without treatment)[64,65]. Such "false-positive" results could be caused by *M.tb* in FNABs that are not culturable using conventional methods like MGIT960. For example, in animal models, *M.tb* DNA in lymph nodes is detectable during re-activation of TB, despite no pathological evidence of disease and no culturability. *M.tb* is hypothesised to then disseminate throughout the body from the lymph node[66]. Moreover, we observed no correlation in bacterial load measured using between Ultra and culture, further supporting the presence of *M.tb* DNA in the absence of culturability.

Critically, if Ultra trace results were excluded or reclassified as an analytical approach to elevate specificity, Ultra would lose sensitivity benefits versus Xpert, however, this sensitivity loss was less for the former strategy than the latter; suggesting exclusion is the preferred strategy for handling trace results if used clinically or in research.

When routine and study Ultra concordance were analysed, study Ultras had higher yield. This may be due to specimen processing (e.g., more sample reagent is used for routine Ultras compared to study Ultras) or cartridge version differences but is overall indicative of an area to improve the diagnosis of TBL within the programme.

Few studies examined Ultra on urine[32-34] and none in patients investigated for TBL. Urine-Ultra may obviate the need for invasive sampling (and hence referral to a specialised facility, and associated costs and delays). Despite concentration[35], low yield and sensitivity were

observed for urine-Ultra, suggesting it could marginally reduce FNAB collection (approximately 1/10). Such a strategy is undermined by elevated non-actionable result rates and cost effectiveness, including the number-needed-to-test, would require prospective investigation and modelling, however, we expect the utility of such an approach to be further enhanced with better urine tests[67].

These results have strengths and limitations. Our study was pragmatic and routine culture not always done and, although our MRS included cytology, multiple cultures (including on specimens from other anatomical sites) may improve specificity estimates. Furthermore, multiple FNAB passes were done to obtain adequate volumes that could have introduced sampling variation, however, FNABs were collected using a standardised protocol by a single health worker.

In conclusion, in a routine clinical setting in patients with presumptive TBL, Ultra detects more TBL than Xpert and would result in more people placed on treatment. This is driven by the added benefit of trace results. Furthermore, programmatic Ultra testing can be optimised on the diagnostic laboratory front, as study Ultra had better performance. Urine-Ultra could reduce invasive sampling and associate delays but there remains a need for better urine-based tests for TBL. We recommend that a positive FNAB Ultra result be used to initiate treatment, however, patients with a negative Ultra still require confirmatory testing and many patients with a trace-positive result will be culture-negative. Our study supports Ultra's use for TBL diagnosis.

Acknowledgments

The authors thank the National Health Laboratory Services, Tygerberg Hospital, Cape Town, South Africa. The authors also thank Ruth Wilson, Lucille Cupido, Roxanne Higgit, Selisha Naidoo and Zaida Palmer. Research reported in this publication was supported by the South African Medical Research Council. The content is the solely the responsibility of the authors and does not necessarily represent the official views of the South African Medical Research Council. The work was funded by the South African Medical Research Council, Stellenbosch University Faculty of Health Sciences, and the National Research Foundation. GT acknowledges funding from the EDCTP2 programme supported by the European Union (grant SF1401, OPTIMAL DIAGNOSIS) and the National Institute of Allergy and Infection Diseases of the National Institutes of Health (U01AI152087). Cepheid donated cartridges but did not have a role in study design or result interpretation.

Figure legends

Figure 1: Specimen collection and diagnostic testing in participants with presumptive TB lymphadenitis. Abbreviations: FNAB, fine needle aspirate biopsy; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Figure 2: Overview of different FNAB-based test results. Tests done as part of the routine diagnostic algorithm (Xpert later replaced by Ultra, cytology, and culture) and the study (Ultra) are shown. Study Ultra detected TB in most culture-positive FNABs and some culture-negative FNABs. Italicised text indicates programmatic testing (programmatic algorithm adherence imperfect). Data are n/N (%). Abbreviations: RIF, rifampicin; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Figure 3: Venn diagrams showing positive results from different FNAB tests (after the 104th participant, Ultra was routinely done instead of Xpert) and urine-Ultra. **(A)** Study Ultra, routine Xpert, culture and cytology results in 59 patients. Study Ultra was positive in seven FNABs undetected by routine Xpert. **(B)** Routine Ultra results relative to Study Ultra, routine Ultra, culture, and cytology in 19 patients. Study Ultra was exclusively positive in 36% (7/19) FNABs not detected by routine Ultra, culture and cytology, and had the highest yield. **(C)** Urine-Ultra results relative to FNAB study Ultra and the MRS in 57 HIV-positive patients (Urine-Ultra negative in all HIV-negatives). Urine-Ultra detects less TBL than FNAB study Ultra but could obviate the need for TB diagnostic FNABs in some patients. Data are n/N (%). Abbreviations: FNAB, fine needle aspirate biopsy; MRS, microbiological reference standard; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Figure 1

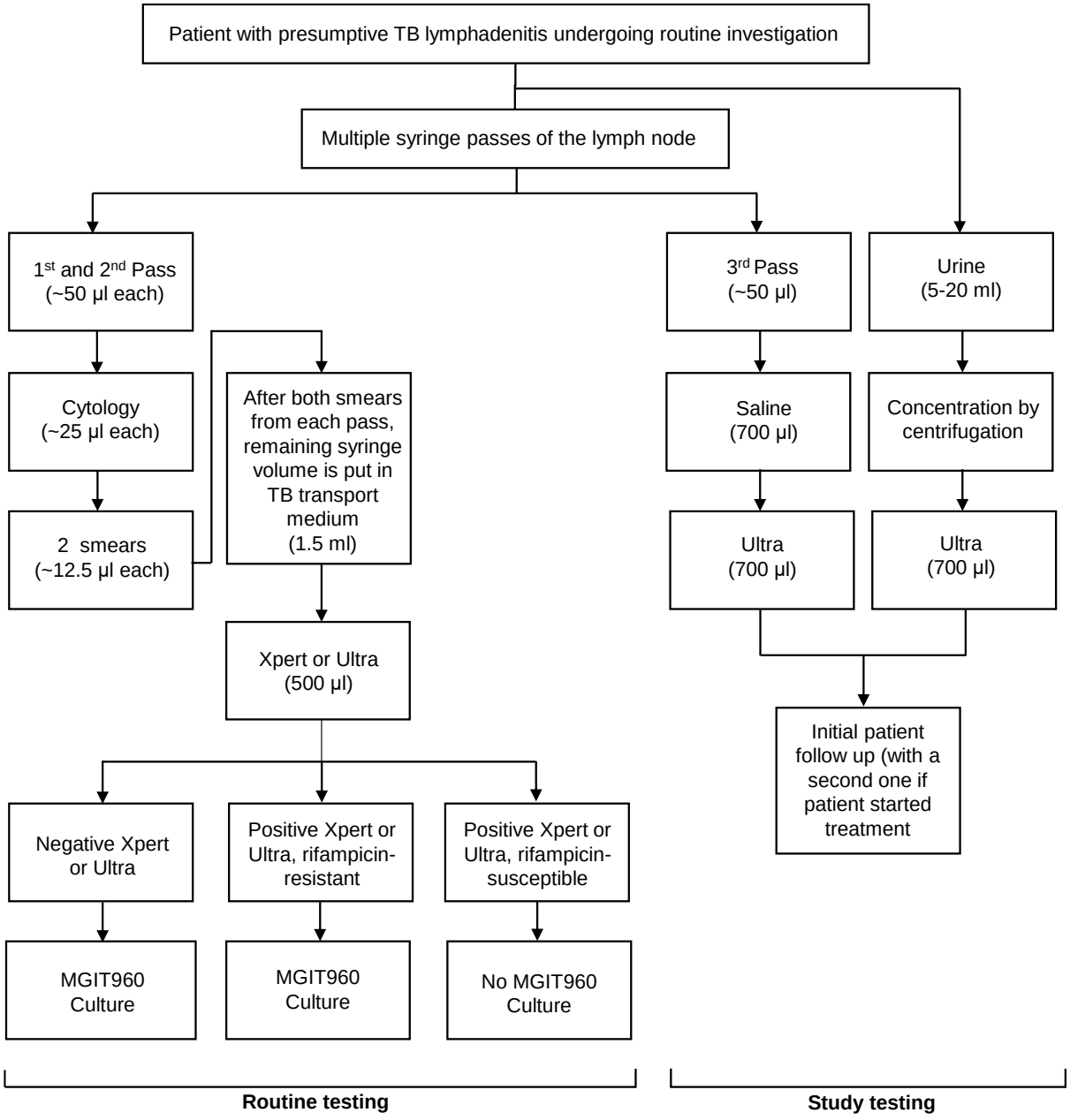
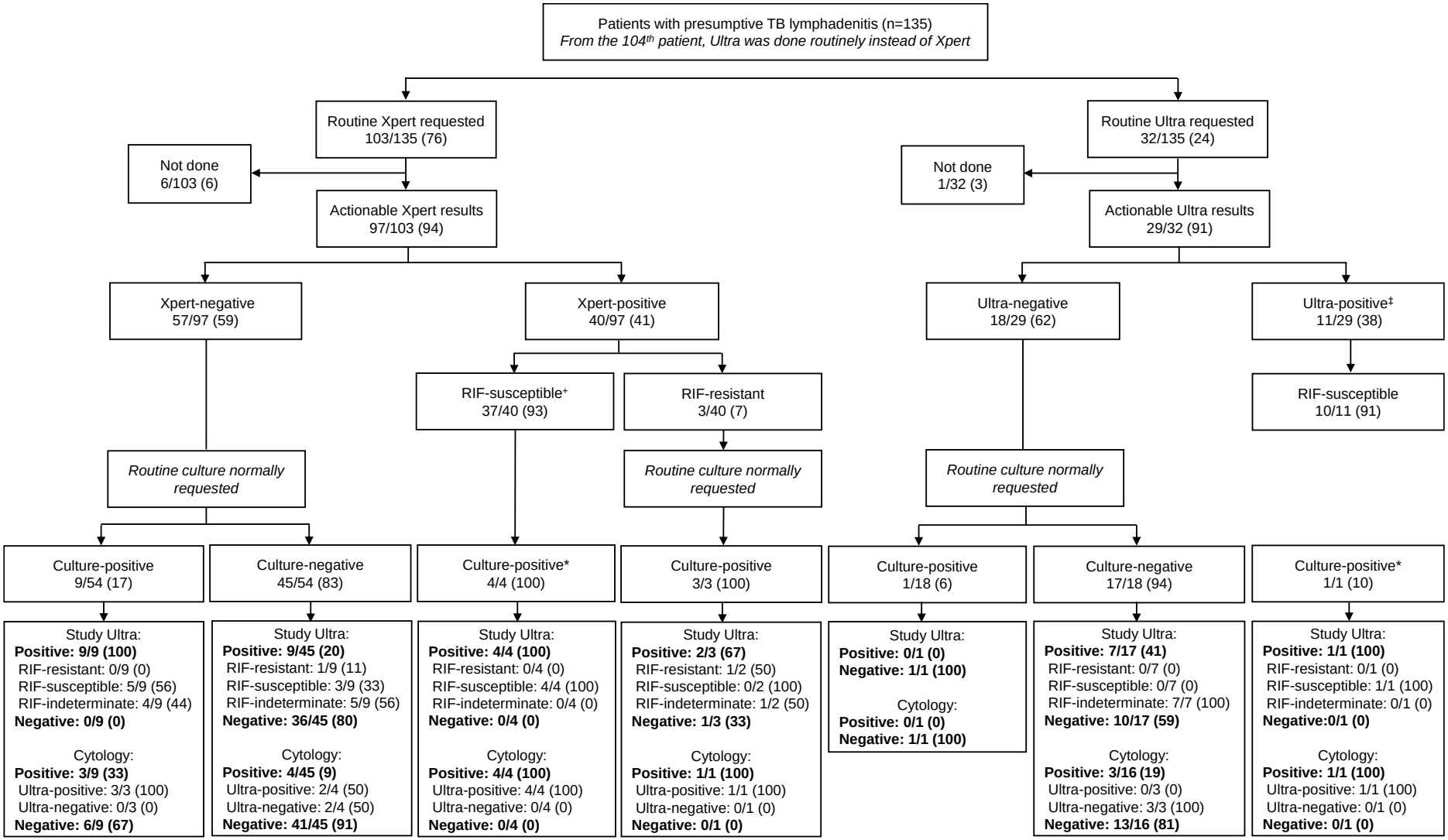


Figure 2



[+]One routine Xpert-positive, rifampicin (RIF)-susceptible patient had a contaminated culture but was study Ultra-positive, RIF-resistant and 32 routine Xpert-positive, rifampicin (RIF)-susceptible patients had no culture per the **Figure 1** algorithm**.**
[ǂ]One routine Ultra was trace-positive RIF-indeterminate.
[*]Culture not normally requested per the routine algorithm.
Ultra results under cytology subheadings (in the last row of boxes) are routine not study Ultras.
Missing data: In patients with a routine Xpert-negative result, one had a contaminated culture and two were culture not done. Two routine Ultras were non-actionable. Three FNABs did not have cytology done.

Figure 3

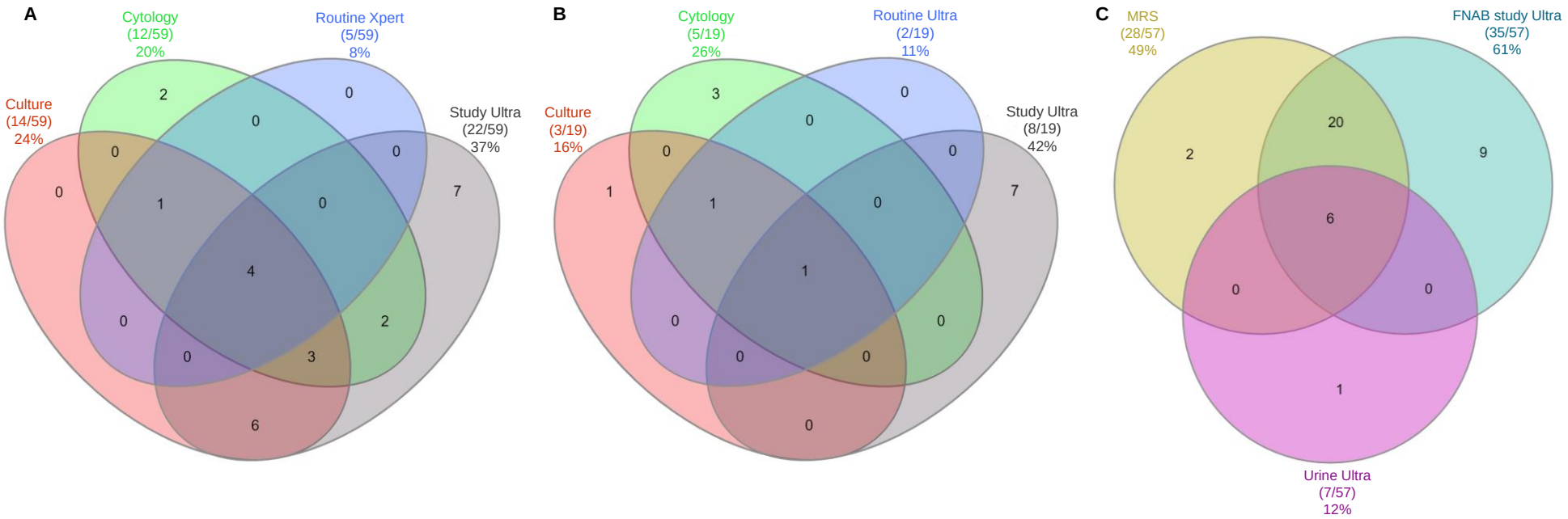


Table 1: Demographic and clinical characteristics by microbiological reference standard status. Definite TBs were more likely to be younger, have an involved neck or breast lymph node (vs. another anatomical site) and, if HIV-positive, a lower CD4 count than non-TBs. Data are n (%) or median (IQR).

	Overall (n=135)	**Definite-TB** (n=59)	**Non-TB** (n=75)
Demographics			
Age (years)	36 (29-46.5)	34 (27-41)	39 (31.5-47.5) **p=0.019**
Female	72/135 (53)	30/59 (51)	42/75 (56) p=0.553
Clinical characteristics			
HIV-positive	77/133 (58)	35/58 (60)	41/74 (55) p=0.569
CD4 count (cells/µl)	183 (66-304)	147 (43-281)	219 (156-358) **p=0.012**
Previous TB	42/135 (31)	19/59 (32)	22/75 (29) p=0.720
Pulmonary TB	38/42 (90)	17/59 (29)	20/75 (27) p=0.783
Extrapulmonary TB	4/42 (10)	2/59 (3)	2/75 (3) p=0.807
Involved site			
Neck	92/134 (67)	53/59 (90)	39/75 (52) **p<0.001**
Thorax	16/134 (12)	4/59 (7)	12/75 (16) p=0.102
Breast	9/134 (7)	0/59 (0)	9/75 (12) **p=0.006**
Other	17/134 (13)	2/59 (3)	15/75 (20) **p=0.004**

Missing data: HIV, two; CD4, four; lymph node site, one.
One patient was unclassifiable based on case definitions.
"Other" sites included arm (n=3), leg (n=3), groin (n=7), and head (n=4).

Table 2: Diagnostic accuracy analyses (non-head-to-head above, head-to-head below) of routine Xpert and study Ultra on FNABs using a MRS (culture and cytology) for *Mycobacterium tuberculosis* complex DNA detection stratified by HIV status. Study Ultra has improved sensitivity compared to routine Xpert but lower specificity. The relative performances of Xpert and Ultra had similar patterns by HIV status and versus the eMRS or CRS (**Supplementary Table 2**). Data are %, 95% CI, and n/N.

	Non-head-to-head											
	All patients				**HIV-negative**				**HIV-positive**			
	n=96				**n=36/96 (38)**				**n=60/96 (62)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert	73 (58, 85) 35/48	92 (80, 98) 44/48	90 (76, 97) 35/39	77 (64, 87) 44/57	65 (38, 86) 11/17	89 (67, 99) 17/19	85 (55, 98) 11/13	74 (52, 90) 17/23	77 (59, 90) 24/31 p=0.343*	93 (77, 99) 27/29 p=0.656*	92 (75, 99) 24/26 p=0.455*	79 (62, 91) 27/34 p=0.627*
	n=130				**n=55/128 (43)**				**n=73/128 (47)**			
Ultra	85 (73, 93) 51/60 p=0.121‡	69 (56, 79) 48/70 **p=0.003‡**	70 (58, 80) 51/73 **p=0.018‡**	84 (72, 93) 48/57 p=0.343‡	76 (55, 91) 19/25 p=0.427‡	70 (51, 85) 21/30 p=0.111‡	68 (48, 84) 19/28 p=0.260‡	78 (58, 91) 21/27 p=0.750‡	91 (76, 98) 31/34 p=0.125‡ p=0.109*	67 (50, 81) 26/39 **p=0.009‡** p=0.768*	70 (55, 83) 31/44 **p=0.031‡** p=0.816*	90 (73, 98) 26/29 p=0.267‡ p=0.227*
Δ if traces excluded	-2 (-15, 12) p=0.808§	+15 (1, 30) **p=0.041§**	+13 (-1, 28) p=0.081§	0 (-13, 13) p>0.999§	-3 (-28, 22) p=0.797§	+21 (1, 41) p=0.076§	+21 (-2, 44) p=0.103§	0 (-22, 22) p>0.999§	-1 (-15, 13) p=0.905§	+12 (-8, 32) p=0.253§	+10 (-9, 28) p=0.332§	0 (-16, 16) p>0.999§
Δ if traces reclassified	-10 (-19, -1) **p=0.014§**	+18 (8, 29) **p<0.001§**	+13 (-1, 28) p=0.081§	-4 (-17, 9) p=0.558§	-12 (-29, 5) p=0.083§	+23 (5, 42) **p=0.008§**	+21 (-2, 44) p=0.103§	-2 (-23, 18) p=0.845§	-9 (-21, 4) p=0.083§	+15 (1, 29) **p=0.014§**	+10 (-9, 28) p=0.332§	-6 (-21, 11) p=0.517§
	Head-to-head											
	n=92				**n=35/92 (38)**				**n=57/92 (62)**			
Xpert	72 (57, 84) 33/46	93 (82, 99) 43/46	92 (78, 98) 33/36	77 (64, 87) 43/56	65 (38, 86) 11/17	94 (73, 100) 17/18	92 (62, 100) 11/12	74 (52, 90) 17/23	76 (56,96) 22/29 p=0.417*	93 (76, 99) 26/28 p=0.832*	92 (73, 99) 22/24 p>0.999*	79 (61, 91) 26/33 p=0.671*
Ultra	91 (79, 98) 42/46 **p=0.016‡**	76 (61, 87) 35/46 **p=0.020‡**	79 (66, 89) 42/53 p=0.114‡	90 (76, 97) 35/39 p=0.105‡	82 (57, 96) 14/17 p=0.244‡	72 (47, 90) 13/18 p=0.074‡	74 (49, 91) 14/19 p=0.217‡	81 (54, 96) 13/16 p=0.593‡	97 (82, 100) 28/29 **p=0.022‡** p=0.099*	79 (59, 92) 22/28 p=0.127‡ p=0.622*	82 (65, 93) 28/34 p=0.311‡ p=0.455*	96 (78, 100) 22/23 p=0.076‡ p=0.145*
Δ if traces excluded	-1 (-17, 11) p=0.836§	+7 (-9, 24) p=0.400§	+5 (-11, 20) p=0.576§	0 (-13, 13) p>0.999§	-3 (=32, 24) p=0.791§	+14 (-12, 41) p=0.321§	+11 (-17, 39) p=0.463§	0 (-27, 27) p>0.999§	1 (-10, 10) p=0.937§	+3 (-18, 24) p=0.787§	+1 (-18, 19) p=0.917§	0 (-12, 12) p>0.999§
Δ if traces reclassified	-13 (-25, 1) **p=0.014§**	+9 (-2, 19) **p=0.046§**	+5 (-11, 20) p=0.576§	-10 (-25, 5) p=0.196§	-17 (-42, 6) p=0.083§	+17 (-6, 39) p=0.083§	+11 (-17, 39) p=0.462§	-8 (-35, 18) p=0.542§	-11 (-25, 4) p=0.083§	+3 (-7, 14) p=0.317§	+1 (-18, 19) p=0.917§	-11 (-26, 5) p=0.219§

Missing data in the non-head-to-head table: Unclassifiable Ultra, n=1; non-actionable Ultras, n=4; HIV, n=2.

Within column p-values: ‡Xpert vs. Ultra within an analysis (non-head-to-head, head-to-head) in patients of the same HIV status (overall, negative, positive).

Within row p-values: *HIV-negative vs. HIV-positive within an analysis (non-head-to-head, head-to-head).

Abbreviations: CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; FNABs, fine needle aspirate biopsies; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Table 3: Study and routine Ultra concordance in patients with both tests done on FNABs. More patients were positive by study Ultra (55%) compared to routine Ultra (35%), corresponding to a 20% incremental yield. Study Ultra had no non-actionable results (column not shown).

		Study Ultra		
		Positive	**Negative**	**Total**
Routine Ultra	**Positive**	10	1	11
	Negative	7	11	18
	Total	18	13	31
	Non-actionable	1	1	2
	Δ Study Ultra vs. routine Ultra	+20% (95% confidence interval; CI: 0, 42) **p=0.034**		

Non-actionable Ultra results included 'Error' (n=1) and 'No result' (n=1).

Abbreviations: Ultra, Xpert MTB/RIF Ultra; FNABs, fine needle aspirate biopsies.

Table 4: Comparison of patient and microbiology characteristics by whether study Ultra was TP or FP per the MRS. FPs were less likely to have been placed on treatment, had less bacterial load, and were less likely to have been detected by routine Xpert and routine Ultra than TPs. Data are n (%) or median (IQR).

	Ultra TPs (n=51)	**Ultra FPs** (n=22)
Patient characteristics		
HIV-positive	31/51 (61)	13/22 (59) p=0.892
CD4 count (cells/µl)	147.0 (32.00-281.30) (n=30)	208.0 (101.3-286.0) (n=12) p=0.238
Previous TB	18/51 (35)	6/22 (27) p=0.503
¥Patients initiated on TB treatment after 12-week follow-up	44/48 (92)	6/22 (27) **p<0.001**
If on treatment, did the patient report improved health?	43/44 (98)	6/6 (100) p=0.709
Study Ultra result information		
rpoB C_{Tmin}	25.70 (20.20-28.20) (n=45)	25.70 (20.40-29.10) (n=9) p=0.878
IS*6110*/IS*1081* C_T	19.00 (16.40-21.60) (n=51)	24.85 (19.88-28.15) (n=22) **p<0.001**
Trace semi-quantitation category	6/51 (12)	13/22 (59) **p<0.001**
SPC C_T	26.10 (25.10-28.60) (n=51)	25.05 (24.45-24.95) (n=22) **p=0.005**
Routine Xpert or routine Ultra information		
Positive Xpert	31/42 (74)	3/11 (27) **p=0.004**
Positive Ultra	7/7 (100)	3/10 (30) **p=0.004**

Missing data: CD4 count, n=2; patients who were lost to follow-up, n=3; unclassifiable routine Xpert results, n=3. True-positive in routine Xpert era not done, n=1; True-positive in routine Ultra era non-actionable, n=1; False-positive in routine Ultra not done, n=1.

Abbreviations: FP, false-positive; IS*6110*/IS*1081* C_T, cycle threshold value for the Xpert MTB/RIF Ultra IS*6110*/IS*1081* probe; *rpoB* C_{Tmin}, minimum cycle threshold value from the Xpert MTB/RIF (Ultra) *rpoB* probes; TP, true-positive; Ultra, Xpert MTB/RIF Ultra. ¥Study Ultra results were not reported for potential patient management.

Chapter 3

Diagnostic accuracy of Xpert MTB/RIF Ultra on pericardial fluid and urine for tuberculosis pericarditis diagnosis in an HIV-endemic setting

Please note in this chapter, TBP refers to TB pericarditis and PF refers to pericardial fluid (this is different in chapter 4)

Supplementary material is attached as Appendix II

Abstract

Background: Tuberculosis pericarditis (TBP) is a deadly manifestation of extrapulmonary TB (EPTB). There are limited data to support Xpert MTB/RIF Ultra (Ultra) on pericardial fluid and urine tests, especially when compared to established biomarkers.

Methods: Adults programmatically investigated for TBP (n=155) underwent study-administered: 1) Ultra (unconcentrated, concentrated) on pericardial fluid (PF), 2) Ultra and 3) Determine TB-LAM (TB-LAM) on urine and 3), on PF, measurement of unstimulated interferon-γ (IFN-γ). The programme did Xpert and later Ultra on concentrated PF and MGIT960 culture. The primary analysis used a microbiological reference standard (MRS).

Results: Unconcentrated PF Ultra (study) had higher sensitivity than Xpert in people with HIV [84% (95% confidence interval: 69, 94) vs. 63% (46, 78); p=0.037] and, overall, lower specificity [69% (57, 79) vs. 93% (84, 97); p<0.001]. Ultra sensitivity and specificity were lower versus the composite reference standard (CRS). When concentrated PF was tested, Ultra specificity increased [83% (72, 90) vs. 69% (57, 79) for unconcentrated PF; p=0.043], and more non-actionable results occurred [12% (18/152) vs. 6% (4/155); p=0.002]. uIFN-γ (rule-out cut-point of >5.10 pg/ml) had high sensitivity [95% (85, 99)] but suboptimal 50% (39, 61) specificity versus the MRS. Both Urine-Ultra and TB-LAM had a yield of 12% (13/109) but detected TBP cases missed by culture.

Conclusions: PF Ultra detects more TBP than Xpert. PF concentration increases Ultra specificity but also non-actionable results. uIFN-γ on PF does not appear useful. Urine tests could reduce the need for pericardiocentesis for TB diagnosis.

243/250

Background

Tuberculosis (TB) is a leading cause of morbidity and mortality[4]. Extrapulmonary TB (EPTB) accounts for 16% of new cases[5] and TB pericarditis (TBP) is one of the deadliest forms of EPTB, particularly in people living with HIV (PLHIV)[13]. TBP is the biggest cause of pericarditis in Africa, responsible for 70% of pericardial effusions in South Africa[68].

Early TBP diagnosis is essential for improving outcomes[69] but remains challenging. Pericardiocentesis, the procedure to collect pericardial fluid (PF) is, due to its technically complex, invasive, and expensive nature[70] only done at high level referral facilities, making it inaccessible for most people requiring investigation for EPTB who only attend primary care. Additionally, culture takes 4-6 weeks and has suboptimal sensitivity for EPTB[3].

Xpert MTB/RIF (Xpert; Cepheid, USA) is a semi-automated real-time PCR that rapidly detects *Mycobacterium tuberculosis* complex DNA (MTBC) and rifampicin resistance[2,71,72]. A systematic review and meta-analysis of Xpert on PF showed a pooled sensitivity of 68% [95% confidence interval (CI): 58, 76] and pooled specificity of 99% (92, 100) versus culture[73]. More sensitive TBP tests are urgently needed and Xpert MTB-RIF Ultra (Ultra), Xpert's successor, offers enhanced limit-of-detection[2]. While studies have evaluated Ultra on different EPTB specimens[39,40], data on TBP remain especially limited.

PF concentrations of unstimulated interferon-γ (uIFN-γ), a cytokine produced by Th1 cells in response to infections, may be diagnostically useful. A meta-analysis showed uIFN-γ had 97% (87, 99) sensitivity and 99% (74, 100) specificity for TBP[69]. This biomarker has therefore been considered for integration into point-of-care assays, although invasive site-of-disease sampling is still required[38,74]. uIFN-γ in conjunction with Ultra is unevaluated, including in comparison with established biomarkers like ADA.

Diagnosing TBP with an easily accessible fluid like urine or blood could mitigate the need for invasive sampling like pericardiocentesis, which is an expensive risky procedure and improve test yield and access. Determine TB LAM Ag (TB-LAM; Abbott, USA) is a lateral flow assay that detects lipoarabinomannan (LAM), a mycobacterial cell wall component in urine[30]. Like TB-LAM, Ultra on urine can improve diagnostic yield[31], however, is not investigated in patients with presumptive TBP.

We undertook a large diagnostic accuracy evaluation of Xpert and Ultra on PF and Ultra and TB-LAM on urine in patients with presumptive TBP in South Africa, measured alongside biomarkers like uIFN-γ. We hypothesised 1) Ultra would show improved sensitivity compared to Xpert and uIFN-γ in PF, 2) concentrating PF would improve Ultra sensitivity, and 3) urinary TB-LAM and Ultra could mitigate pericardiocentesis.

Methods and materials

Ethics statement

The study was approved by the Stellenbosch University Human Research Ethics Committee (N16/04/050) and Tygerberg General Hospital (TGH; WC_2016RP15_762).

Patient recruitment and sampling

155 inpatients (≥18 years) with an ultrasound-identified pericardial effusion that required pericardiocentesis as part of programmatic care at TGH in Cape Town, South Africa, were prospectively consecutively recruited from 24 November 2016-17 January 2022. Programmatically administered chest X-rays (CXRs) were read by a radiologist. PF, blood, and urine were collected. Due to the lengthy referral process required for patients to undergo pericardiocentesis at TGH and frequent programmatic use of empiric treatment[75], patients on treatment ≤2 weeks were eligible.

Pericardial fluid collection and programmatic testing algorithm

PF and, if clinically indicated, pericardial biopsies were collected and sent for programmatic testing at the National Health Laboratory Service (NHLS). Excess fluid was kept for study testing (**Figure 1**) blinded to programmatic results.

Definitions

Patients were designated as definite- or non-TBP. For the microbiological reference standard (MRS), TBPs were culture-positive on PF or pericardial biopsy, and non-TBPs culture-negative on PF and, if done, pericardial biopsy. The extended microbiological reference standard (eMRS) included tests on fluid other than PF and biopsies. The composite reference standard (CRS) included follow-up treatment information (**Supplementary Table 1**). Actionable results were those that provided clinically useful information (non-actionable results are not positive or negative, defined further in **Supplement**).

Programmatic testing

Pericardial fluid and biopsy testing

Culture: 5-7.5 ml PF or biopsy were NALC-NaOH-decontaminated, centrifuged, resuspended in phosphate buffer, and 500 µl inoculated into a MGIT960 tube. Genotype MTBDR*plus* (Hain Lifesciences, Germany) was done on isolates for speciation and drug susceptibility testing[76].

Xpert and Ultra: Xpert (version 1; Cepheid, USA) was done from 25 January 2017–9 April 2018 and thereafter replaced with Ultra (version 1). Sample reagent (SR) buffer (1.5 ml) was added to 500 µl (3:1) of concentrated PF or biopsy and 2 ml used for Xpert or Ultra[57,58]. The programmatic algorithm for culture and PCR differed if the patient had PF previously collected within the last three months (**Supplementary figure 1**).

Cytochemistry: Adenosine deaminase (ADA) was quantified using the Diazyme Adenosine Deaminase assay (Diazyme Laboratories, USA) and total protein, lactate dehydrogenase (LDH) and albumin quantified using the SYNCHRON (Beckman Coulter, USA) system.

Study testing

Pericardial fluid

Xpert and Ultra: Two 700 µl unconcentrated PF aliquots were each freshly tested with Xpert (started by the study when the programme switched to Ultra) and Ultra (both version 3) using 1.4 ml sample reagent buffer (2:1 volume ratio) as per manufacturer's instructions[57,58]. Volume permitting, concentrated Ultra was done after centrifuging 20 ml fresh PF (1711×*g*, 10 min, room temperature) and removing supernatant until 700 µl remained. The resuspended pellet was treated with 1.4 ml sample reagent (2:1). If non-actionable (see below Definitions section), Xpert and Ultra were repeated once on a new PF-sample reagent mix.

Unstimulated IFN-γ: IFN-γ concentrations (without antigen stimulation) were measured in duplicate using filter-sterilised supernatant (0.45 µm then 0.22 µm filtration; Thermo Scientific, USA) from 1.4 ml of centrifuged (20124×*g*, 15 min, room temperature) PF after one freeze-thaw cycle. A standard curve (acting as positive controls) and negative controls were in each plate tested using the Human IFN-γ ELISAPRO kit (Mabtech, Sweden) according to manufacturer's instruction[77].

Urine-Ultra and TB-LAM

5-20 ml urine stored at -80 ºC for a median of 505 days [interquartile range (IQR): 75-747] or freshly collected were centrifuged (2862×*g*, 15 min, room temperature) and supernatant removed until ~700 µl remained. The pellet was resuspended with 1.4 ml sample reagent buffer and tested by Ultra[58]. 700 µl of unconcentrated urine was tested by Ultra when the concentrated Ultra was positive or non-actionable. ~60 µl unconcentrated urine was tested using TB-LAM[78].

Patient treatment and follow-up

Treatment decisions were programmatic (no study results reported). Patients were telephonically followed-up ≥12 weeks post-recruitment. TB treatment was recorded and, if started, patients self-reported treatment response at follow-up. Patients were lost-to-follow-up

if at least two calls and messages were unsuccessful.

Statistical analysis

Unless otherwise stated, we present results versus the MRS and, for comparisons with other tests, head-to-head results. The final Xpert or Ultra result (i.e., after retesting if initially non-actionable) are included. Proportion and McNemar's tests were done using STATA version 16.0 (StataCorp, USA)[79] and GraphPad Prism version 8.0.1 (GraphPad Software, USA). Euler diagrams were made with InteractiVenn[60]. STARD guidelines were followed[61]. Diagnostic yield was calculated as, of people with a positive test result [study Ultra on PF (unconcentrated or concentrated), study Xpert on PF (unconcentrated or concentrated), programmatic PF culture, programmatic culture on pericardial biopsy, programmatic smear microscopy, culture, concentrated Ultra and concentrated Xpert-all on separate non-site-of-disease fluid, study urine-Ultra (concentrated or unconcentrated), study urine TB-LAM], the proportion positive by one specific test.

Results

Patient characteristics

3% (5/155) patients were MRS-unclassifiable and, of the remainder, 40% (60/150) were TBP and 60% (90/150) non-TBP. People with TBP were more likely to have HIV, fever and night sweats and have started treatment by 12-week follow-up compared to non-TBPs. Furthermore, TBP's PF was more likely to be purulent and have higher ADA, albumin and uIFN-γ concentrations (**Table 1)**.

Xpert and Ultra on PF

Xpert and Ultra results stratified by MRS and whether PF was concentrated are in **Figure 2**.

Rifampicin results

Both unconcentrated and concentrated Ultra correctly identified rifampicin resistance in two MTBDR*plus*-resistant patients (**Supplementary Table 2**). Xpert incorrectly identified one rifampicin-resistant case. Specificities were 50% (1/2), 100% (2/2) and 100% (2/2) for Xpert, unconcentrated Ultra and concentrated Ultra respectively.

Non-actionable results

Unconcentrated Xpert had an initial non-actionable result rate of 7% (10/148) [only includes study Xpert results, programmatic Xpert non-actionable result rates unavailable], which was similar to that for unconcentrated Ultra [6% (4/155)]. After concentration, the Ultra non-actionable result rate increased to 12% (18/152; p=0.002). Re-testing resolved most non-actionables with 80% (8/10), 100% (4/4) and 69% (11/69) becoming actionable for unconc. study Xpert, unconc. Ultra, and conc. Ultra, respectively (**Supplementary Table 4**). When the characteristics of patients and specimens were compared between those who were Ultra actionable or non-actionable (after retesting, if done), non-actionables had lower median uIFN-

γ [0 pg/ml (IQR: 0-649) vs. 906 pg/ml (3-2576); p=0.002] and protein [50 g/L (42-62) vs. 59 g/L (51-66); p=0.016] concentrations (**Supplementary Table 5**).

Sensitivity and specificity of Xpert and Ultra

Overall: When unconcentrated Ultra was compared to Xpert (**Table 2**), sensitivity was [80% (95% confidence interval: 67, 90) vs. 64% (50, 76; p=0.057)] and specificity was [69% (57, 79) vs. 93% (84, 97; p<0.001]. Compared to unconcentrated Ultra, concentrated Ultra had similar sensitivity [85% (73, 94); p=0.449] and increased specificity [83% (72, 90); p=0.043]. Analyses using eMRS had similar conclusions, however, when the CRS was used, sensitivity loss occurred for Xpert [64% (50, 76) vs. 39% (30, 50; p=0.004], unconcentrated Ultra [80% (67, 90) vs. 58% (47, 67; p=0.005] and concentrated Ultra [85% (73, 94) vs. 60% (49, 69; p=0.001] while specificity was similar (**Supplementary Table 8**). Conclusions were unchanged for non-head-to-head comparisons (**Supplementary Tables 7 and 9).**

In people without HIV: Unconcentrated Ultra had similar sensitivity [71% (44, 90) vs. 65% (38, 86); p=0.714] and decreased specificity [64% (49, 78) vs. 93% (82, 99); p=0.001] to Xpert (**Table 2**). When concentrated and unconcentrated Ultras were compared, sensitivity was similar [82% (57, 96) vs. 71% (44, 90); p=0.419] and specificity increased [89% (76, 96) vs. 64% (49, 78); p=0.006]. Conclusions were similar for non-head-to-head comparisons (**Supplementary Table 7)**.

In PLHIV: Unconcentrated Ultra had improved sensitivity [84% (69, 94) vs. 63% (46, 78); p=0.037] and similar specificity [74% (57, 88) vs. 91% (77, 98); p=0.057] to Xpert. When concentrated and unconcentrated Ultras were compared, sensitivity [87% (72, 96) vs. 84% (69, 94); p=0.744] and specificity [74% (57, 88) vs. 74% (57, 88); p>0.999] were similar. Conclusions were unchanged for non-head-to-head comparisons (**Supplementary Table 7)**.

Comparisons of individual tests in people with or without HIV: Xpert sensitivity, specificity and PPV were similar but NPV in PLHIV decreased versus that in without HIV [70% (54, 82)

vs. 88% (75, 95); p=0.034] (**Table 2**). Unconcentrated Ultra only differed in PLHIV compared to HIV-negatives in terms of PPV [78% (62, 89) vs. 43% (24, 63); p=0.003] and concentrated Ultra accuracy showed no differences across HIV statuses.

Trace recategorization strategies: When unconcentrated Ultra traces were excluded, sensitivity was unchanged [-3% (-19, 13); p=0.697] and specificity increased [+18% (5, 31); p=0.010] (**Supplementary Table 6**). When these results were rather reclassified as negative, sensitivity decreased [-20% (-37, 3), p=0.022] and specificity increased [+21% (9, 33), p=0.001]. When concentrated Ultra traces were excluded or reclassified as negative, sensitivity and specificity was unchanged.

Ultra inhibition: Sample processing control (SPC) C_T values showed inhibition in study concentrated Ultras versus study unconcentrated Ultras [26.05 (IQR: 24.70-27.30) vs. 25.30 (24.50-26.20); p<0.001] (**Supplementary Figure 1A**).

Ultra rpo*B* C_{Tmin} and IS*6110*/IS*1081* C_T: Study concentrated Ultra had lower *rpo*B C_{Tmin} compared to unconcentrated Ultra [25.10 (IQR: 20.53-26.10) vs. 27.50 (23.63-30.15); p<0.001] and lower IS*6110*/IS*1081* C_T [18.65 (16.90-21.08) vs. 21.55 (18.60-23.15); p<0.001] values (**Supplementary Figure 1B and 1C)**, demonstrating the effect of concentration.

Relationship with bacterial load: Neither Xpert nor study unconcentrated Ultra *rpo*B C_{Tmin} correlated with culture time-to-positivity (TTP), however, Ultra IS*6110*/IS*1081* C_T had a positive linear correlation. After concentration, both study concentrated Ultra *rpo*B C_{Tmin} and IS*6110*/IS*1081* C_T correlated with TTP (**Supplementary Figure 2A-E**).

Programmatic Ultras vs. study Ultras: When study concentrated and unconcentrated Ultra positive results were separately compared to programmatic concentrated Ultras, results were similar (**Supplementary Table 10A and B**). Study unconcentrated Ultra results and study concentrated Ultra results showed no difference (**Supplementary Table 10C)**.

uIFN-γ on PF

uIFN-γ sensitivity and specificity

Optimal cut-points: Of the 155 patients enrolled, 2% (3/155) did not have uIFN-γ done. Optimal rule-in, rule-out and Youden's index cut-point concentrations for uIFN-γ on PF to maximise diagnostic accuracy were determined using the ROC curve shown in **Figure 4A**. The area under the ROC curve (AUROC) for all patients was 0.76 [95% confidence interval (CI): 0.69, 0.84]. We prioritised sensitivity for diagnostic accuracy analyses (**Table 2**) by using a cut-point of >5.10 pg/ml, which corresponds to a sensitivity of 95% (88, 99) and 53% (44, 62) specificity. Specificity being prioritised and Youden's index can be seen in the supplement. The AUROC and cut-points for PLHIV and people without HIV (**Figure 4B** and **4C)** can be seen in supplementary results (**Supplementary Table 3**).

Diagnostic accuracy (head-to-head): When unconcentrated Ultra and uIFN-γ were compared, uIFN-γ (5.10 pg/ml) had higher sensitivity [95% (85, 99) vs. 80% (67, 90); p=0.022] and lower specificity [50% (39, 61) vs. 69% (57, 79); p=0.016] (**Table 2**). When unconcentrated Ultra and uIFN-γ (5.1pg/ml) were compared in PLHIV, uIFN-γ had higher sensitivity [100% (79, 100) vs. 69% (41, 89); p=0.015] and similar specificity [61% (45, 76) vs. 64% (48, 78); p=0.826]. In people without HIV, uIFN-γ had similar sensitivity [92% (79, 98) vs. 85% (69, 94); p=0.919] and decreased specificity [37% (21, 55) vs. 74% (57, 88); p=0.002]. Conclusions were unchanged for non-head-to-head comparisons (**Supplementary Table 7**). When PLHIV were compared to patients without HIV, PLHIV had similar sensitivity, decreased specificity [37% (21, 55) vs. 61% (45, 76); p=0.032], similar PPV and decreased NPV [81% (54, 96) vs. 100% (87, 100); p=0.020]. For uIFN-γ, PLHIV had similar sensitivity, decreased specificity [37% (21, 55) vs. 60% (44, 74); p=0.043], similar PPV and decreased NPV [81% (54, 96) vs. 100% (87, 100); p=0.020].

When actionable results were compared for: 1) MRS, 2) Xpert, 3) unconcentrated Ultra, 4) concentrated Ultra and 5) uIFN-γ **(**rule-out cut-off 5.1 pg/ml**)**, 41% (55/135), 30% (41/135), 51% (69/135), 45% (61/135), 69% (93/135) were positive by each test (**Figure 3A**); uIFN-γ had the highest yield followed by unconcentrated Ultra.

Urine results

Tests on urine detected TBP in some cases missed by tests on PF

Of concentrated urine specimens tested with Ultra (n=99), 15% (15/99) were non-actionable and 100% (15/15) resolved to actionable when unconcentrated urine was tested (20% (3/15) unconcentrated urines became Ultra-positive). When Urine-Ultra was compared to TB-LAM using the MRS (*Overall, head-to-head*; n=76) (**Table 3**), Ultra had similar sensitivity [25% (95% confidence interval: 11, 44) vs. 29% (13, 49); p=0.763] and similar specificity [92% (80, 98) vs. 90% (77, 97); p=0.726]. Data was unchanged after stratification by HIV status. Excluding TB-LAM, Ultra detected 13% (4/32) TB not detected by the MRS (**Figure 3B**). When Urine-Ultra was excluded, TB-LAM detected 15% (5/33) TBP cases missed by the MRS (**Figure 3B**).

When actionable results were compared for the 1) MRS on PF, 2) unconcentrated and concentrated Ultra on urine, 3) TB-LAM on urine, 37% (28/76), 14% (11/76), 17% (13/76) were positive by each test (**Figure 3B**; both tests on urine had a low yield**)**.

When actionable results were compared for the 1) MRS on PF, 2) unconcentrated Ultra on PF, 3) concentrated Ultra on PF, 4) unconcentrated and concentrated Ultra on urine and 5) TB-LAM on urine, 38% (28/73), 47% (34/37), 42% (31/73), 15% (11/73), 18% (13/73) were positive by each test (**Figure 3C**; unconcentrated Ultra on PF followed by concentrated Ultra on PF had the highest yield).

Diagnostic Yield

Overall (patients who did or did not have the test attempted)

Study Ultra (unconc. and conc.) on PF had the highest yield 83% (91/109) followed by culture on PF 52% (57/109), Xpert (unconc. and conc.) on PF 41% (45/109) and culture on non-site-of-disease fluid 30% (33/109) (**Supplementary Table 14**). Culture on pericardial biopsy 13% (15/109), most tests on non-site-of-disease fluid [smear microscopy 5% (5/109), programmatic Ultra 12% (13/109), programmatic Xpert 9% (10/109)] and urine tests [Ultra (unconc. and conc.) 12% (13/109), TB-LAM 12% (13/109)] had low yields respectively. Culture on PF and non-site-disease fluid and Urine-Ultra (unconc. and conc.) had higher yields in PLHIV than in people without HIV [culture on PF: 65% (40/52) vs. 36% (17/47); p=0.003], [culture on non-site-of-disease fluid: 39% (24/62) vs. 19% (9/47); p=0.028] and [Urine-Ultra (unconc. and conc): 19% (12/62) vs. 2% (1/47); p=0.006].

Overall (in patients who had the test attempted)

Smear microscopy on non-site-of-disease fluid had the highest yield 100% (5/5), followed by culture on pericardial biopsy 63% (15/24) and study Ultra (unconc. and conc.) on PF 59% (91/155) (**Supplementary Table 14**). Culture on PF had a yield of 38% (57/151), while Xpert on PF and culture, programmatic Ultra and programmatic Xpert on non-site-of-disease fluid had yields of 30% (45/152), 37% (33/89), 23% (13/56) and 26% (10/39). Urine tests had low yields; Ultra (unconc. and conc.) 13% (13/99) and TB-LAM 16% (13/79).

Comparison of unconcentrated Ultra true-positives and false-positives per the MRS

True-positives were more likely to be HIV-positive than false-positives [77% (36/47) vs. 30% (9/30), p<0.001] but if HIV-positive, more false-positives were on ART compared to true-positives [63% (5/8) vs. 25% (9/36), p<0.039] (**Supplementary Table 13**). More true-positives started TB treatment after being followed up [98% (46/47) vs. 50% (14/28), p<0.001] and were

more likely to have a fever [50% (22/44) vs. 21% (6/29), p=0.012] or experience night sweats [62% (28/45) vs. 29% (8/28), p=0.005] than false-positives. False-positives had higher *rpoB* $C_T$min and IS*6110*/IS*1081* C_T than true-positives [*rpoB* C_{Tmin} 29.50 (28.90-35.50) vs. 27.40 (22.14-29.75); p=0.008] and [IS*6110*/IS*1081* C_T 25.50 (23.23-27.25) vs. 20.85 (18.33-23.15); p<0.001] and, a greater proportion of false-positives were hence trace-positive [70% (21/30) vs. 23% (11/48); p<0.001]. More true-positives were concordant with positive Xperts and programmatic Ultras than false-positives [Xpert-positives: 79% (37/47) vs. 17% (5/30); p<0.001] and [programmatic Ultra-positives: 100% (13/13) vs. 27% (4/15); p<0.001]. True-positives had higher uIFN-γ [1495 pg/ml (691-2883) vs. 3.65 (0-1720); p<0.001] concentrations compared to false-positives. More true-positives were eMRS-positive [100% (48/48) vs. 17% (5/30); p<0.001] and CRS-positive [100% (48/48) vs. 47% (14/30); p<0.001] compared to false-positives. The characteristics of false-positives per patient information is in **Supplementary Table 11**.

Patient treatment status at follow-up

96% (149/155) of patients were followed-up [median (IQR: 24 (15-46) weeks since recruitment] and 76% (113/149) of those had initiated TB treatment. Of these, 50% (57/113) were classified as definite TB and 47% (53/113) as non-TB per the MRS [3% (3/113) were unclassifiable]. Of the definite TBs, 81% (46/57) were unconcentrated Ultra-positive and for non-TBs, 26% (14/53) were unconcentrated Ultra-positive. Xpert, concentrated Ultra and uIFN-γ can be seen in Supplementary results. Regarding the clinical status in patients who started treatment, 93% (103/111) reported treatment completion and, of these, 94% (95/101) reported feeling clinically well. 19% (28/145) patients were documented to have died, of whom 11% (3/28) who were not placed on treatment were exclusively unconcentrated Ultra-positive.

Discussion

Our key findings are: 1) in PLHIV Ultra on unconcentrated PF had, vs. Xpert, had increased sensitivity and decreased specificity overall; 2) Exclusion of unconcentrated Ultra trace results improved specificity without large sensitivity decrements (unlike reclassifying traces to negative); 3) Ultra on concentrated PF showed improved specificity and similar sensitivity vs. unconcentrated PF, but concentrated Ultra came with increased non-actionable results; 4) uIFN-γ has high sensitivity on PF but has moderate specificity; 5) Ultra and TB-LAM on urine detected few TBP cases but could reduce the proportion of patients who require pericardiocentesis for TBP diagnosis by 4%. These data demonstrate unconcentrated Ultra can increase the yield of TBP diagnosis.

In PLHIV, unconcentrated Ultra on PF had increased sensitivity vs. Xpert compared to the MRS, suggesting that Ultra can be used as a rapid initial test for TBP diagnosis[3]. This increased sensitivity came with an overall loss in specificity, however, with unconcentrated Ultra resulting in approximately 3 in 10 MRS-negative, unconcentrated Ultra-positive TBP cases (this was unchanged when the eMRS and CRS was used). This is consistent with previous studies that showed when directly compared to Xpert, Ultra had improved sensitivity and decreased specificity in fine needle aspirates[80], a combination of EPTB specimens[40] and sputum[81]. Loss in specificity was not associated with previous TB as seen in pulmonary TB[53], but could be due to the imperfect nature of reference standards in EPTB specimens[24,49] as approximately 50% study MRS-negative unconcentrated Ultra-positives (‘false-positives’) were CRS-positive.

Critically, if unconcentrated Ultra trace results were excluded as an analytical approach, specificity increased. Moreover, if trace results were reclassified as negative, specificity improvement came with a loss in sensitivity. This improvement in specificity was also seen in

sputum[53] when Ultra trace results were reclassified as negative, and in one EPTB study on fine needle aspirate biopsies[80] when Ultra trace results were reclassified as negative or excluded. In PF trace exclusion is the superior strategy for handling trace results if used clinically or in research.

Concentrated Ultra notably showed improved specificity compared to unconcentrated Ultra. This has not been observed in EPTB specimens before. We speculate that concentrated Ultra's improved specificity might be due to increased PCR inhibitors (high SPC C_T) preventing possible concentrated Ultra false-positives from emerging as they now are true-negatives, thereby increasing specificity. Additionally, concentration of PF does come with an increase in non-actionable concentrated results compared to unconcentrated Ultra (concentrating less PF might reduce non-actionable results). The benefit of PF concentration might thus be negated by increased non-actionable results which is noteworthy as concentrated PF is used for programmatic Ultras in South Africa.

uIFN-γ has high sensitivity when the rule-out cut-off of 5.1 pg/ml was used, but falsely identifies 50% TBP cases. This is dissimilar to a systematic review that showed both high sensitivity and specificity in four studies which all had an area under the curve (AUC) above 0.90[69], unlike our study that had an AUC of 0.76. This could be due to the use of different IFN-γ assays (as it is unknown whether all studies included in the review used an unstimulated assay), a few studies had small cohorts, and all used composite reference standards which included TB tests with low specificity. The potential use of uIFN-γ is also hindered by the laboratory labour and instrumentation required.

More sensitive tests that ideally use non-invasive specimens than PF are still needed. However the low sensitivity of urine-Ultra and TB-LAM in this cohort mirrors that in a prior study of patients with presumptive TBP[82] and the low positivity rate demonstrated in patients with TB-

meningitis[33]. Nonetheless Urine-Ultra identified TB missed by culture, suggesting universal urine testing could reduce the number of patients undergoing pericardiocentesis in a subset of patients. Lot variability has impacted ongoing evaluations of Fujifilm SILVAMP TB LAM[83]; higher sensitivity 3rd-generation LAM assays are eagerly awaited.

Our data represent one of the largest cross-sectional studies evaluating molecular, microbiological and biomarker tests for TBP diagnosis. Pragmatic limitations of our study include that programmatic Xpert was not always done, therefore study Xpert (which was unconcentrated) was done. We were therefore unable to measure head-to-head the change in sensitivity and specificity associated with using concentrated rather than unconcentrated PF, as we were able to do with Ultra. PF is decontaminated at the NHLS (where programmatic tests are done) and further diluted with more sample reagent (SR) buffer for Xpert than recommended[58], which may underestimate Xpert sensitivity. Moreover, the sensitivity of TB culture could be underestimated as a subset of patients were on empirical treatment, but we contend that the effect would be minimal as the median time on treatment was one day.

In conclusion, Ultra (with the inclusion of trace results) confirms TBP in more people with presumptive TBP than Xpert, but this comes with a loss in specificity. Concentrating PF improves Ultra specificity, but the increased non-actionable results may negate this benefit. If laboratories have sufficient PF and capacity for re-testing however, we recommend doing an Ultra with concentrated PF due to Ultra's increased specificity. The high sensitivity of uIFN-γ is offset by poor specificity, and high costs. Urine testing may reduce the need for invasive sampling in a small subset of patients. We suggest that all patients with presumptive TBP first receive a urine-based test; if positive, TB treatment should commence and if negative, an unconcentrated Ultra on PF should be done. Our data support the use of unconcentrated Ultra

with the inclusion of trace Ultra results for TBP diagnosis in programmatic practice, while emphasising the urgent need for higher accuracy tests on non-invasively collected specimens.

Acknowledgments

The authors thank the study participants, National Health Laboratory Services, Tygerberg Hospital, Dawn Davis, Ruth Wilson, Nandi Niemand, Roxanne Higgit, Selisha Naidoo, Dr Megan Hendrikse and Zaida Palmer. Research reported in this publication was supported by the South African Medical Research Council. GT acknowledges funding from the EDCTP2 programme supported by the European Union (RIA2018D-2509, PreFIT; RIA2018D-2493, SeroSelectTB; RIA2020I-3305, CAGE-TB) and the National Institutes of Health (D43TW010350; U01AI152087; U54EB027049; R01AI136894). CN is supported by the Fogarty International Center of the National Institutes of Health (K43TW012302). The degree from which this study emanated was made possible through funding by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the SAMRC Internship Scholarship Programme. We acknowledge the support of the South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. The content is solely the responsibility of the authors and does not necessarily represent the official views of any funders. Cepheid donated cartridges but had no role in study design or analysis.

Figure legends

Figure 1: Specimen collection and diagnostic procedures done in participants with suspected TB pericarditis. Abbreviations: conc., concentrated; PF, pericardial fluid; TB, tuberculosis; TB-LAM, Determine TB-LAM; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF; uIFN-γ, unstimulated interferon-γ.

Figure 2: Summary of test results done programmatically (culture, concentrated Xpert, concentrated Ultra) and by the study (unconcentrated and concentrated Ultra, unconcentrated Xpert when the programme did not do Xpert) are shown, from either PF or biopsies. Unconcentrated Ultra detected TBP in many culture-negative specimens missed by concentrated Ultra and Xpert. Concentration did not increase Ultra positivity. Data are n/N (%). Abbreviations: conc., concentrated; PF, pericardial fluid; RIF, rifampicin; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

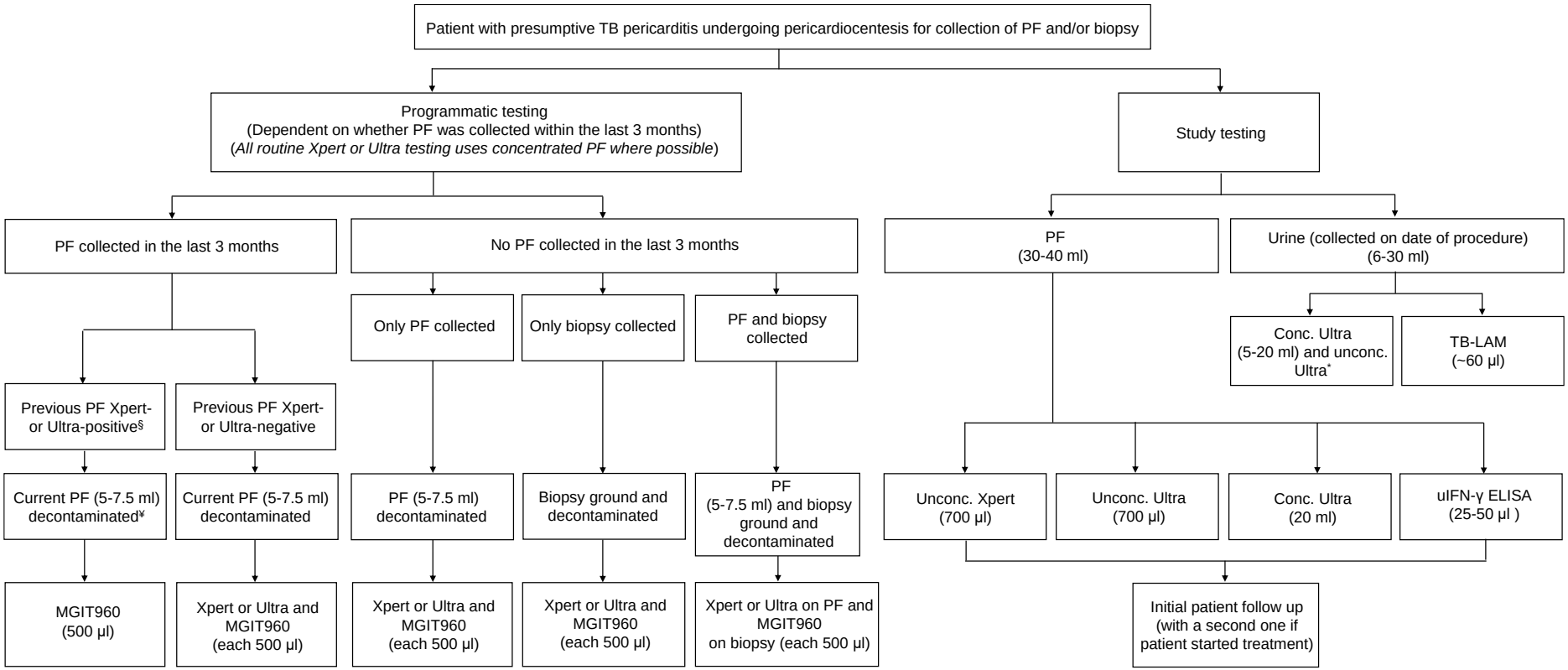
Figure 3: Euler diagrams showing positive results from tests done on PF and/or urine versus the MRS. **(A)** Xpert, study Ultra (unconcentrated and concentrated) and uIFN-γ (cut-off 5.1 pg/ml) on PF. Both PF study unconcentrated Ultra and uIFN-γ were positive in patients not detected by Xpert, study concentrated Ultra, and the MRS. **(B)** MRS, urine study Ultra (unconcentrated and concentrated) and urine TB-LAM results irrespective of HIV status. Urinary TB-LAM and Ultra detected patients the MRS missed. **(C)** MRS, study unconcentrated Ultra on PF, study concentrated Ultra on PF and, on urine, study unconcentrated and concentrated Ultra and TB-LAM (irrespective of HIV status). PF study unconcentrated Ultra followed by study concentrated Ultra on PF detected patients missed by the MRS and urine-based tests. Urine tests detected TB in cases missed by PF tests. Data are n/N (%). Abbreviations: conc., concentrated; MRS, microbiological reference standard; PF, pericardial

fluid; TB-LAM, Determine TB-LAM; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF; uIFN-γ, unstimulated interferon-γ.

Figure 4: Receiver operator characteristic (ROC) curves of uIFN-γ **(A-C)**, ADA **(D-F)** and albumin (**G-I**) on PF from all, HIV-positive, or HIV-negative patients. Values are AUCs with 95% CIs (dashed lines). Abbreviations: ADA, adenosine deaminase; AUC, area under the receiver operating characteristic curve; CIs, confidence intervals; ROC, receiver operator characteristics; uIFN-γ, unstimulated interferon-γ.

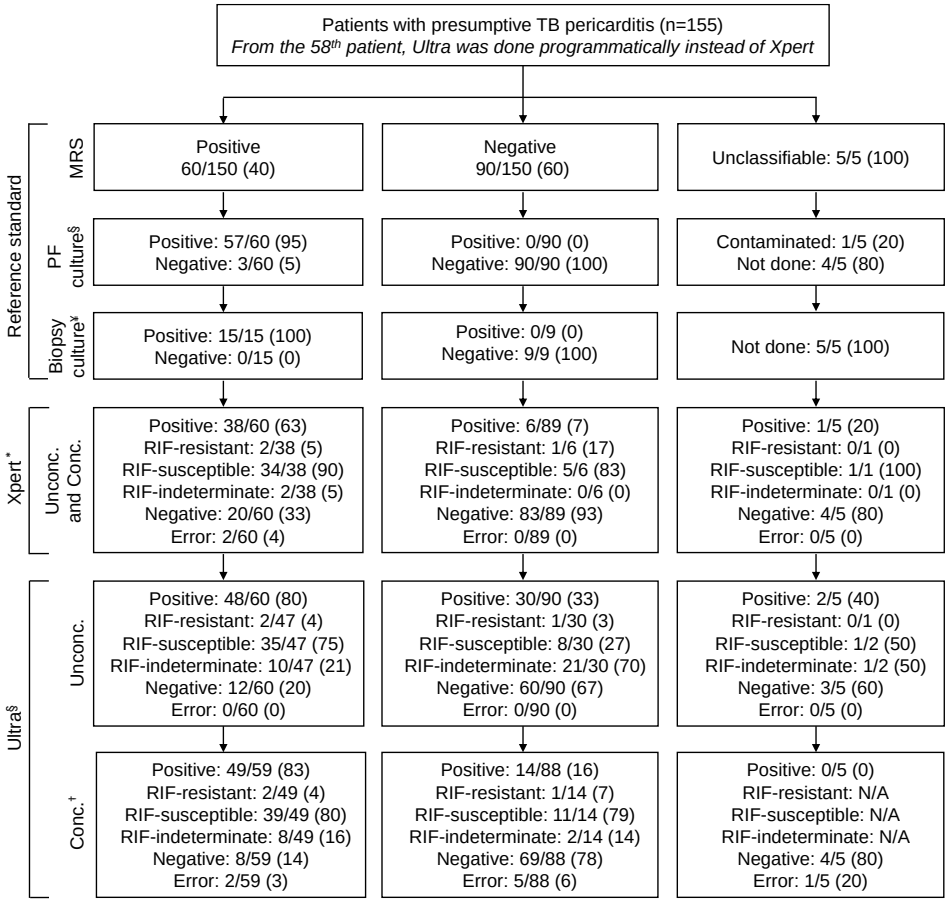
Figure 5: Sensitivity and specificity of tests on PF and urine compared to the MRS, eMRS and CRS. Ultra and uIFN-γ on PF had high sensitivity, and Ultra had lower specificity compared to Xpert. Tests on urine had low sensitivity. Abbreviations: CI, confidence interval; CRS, composite reference standard; conc., concentrated; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; PF, pericardial fluid; uIFN-γ, unstimulated interferon-γ, Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Figure 1



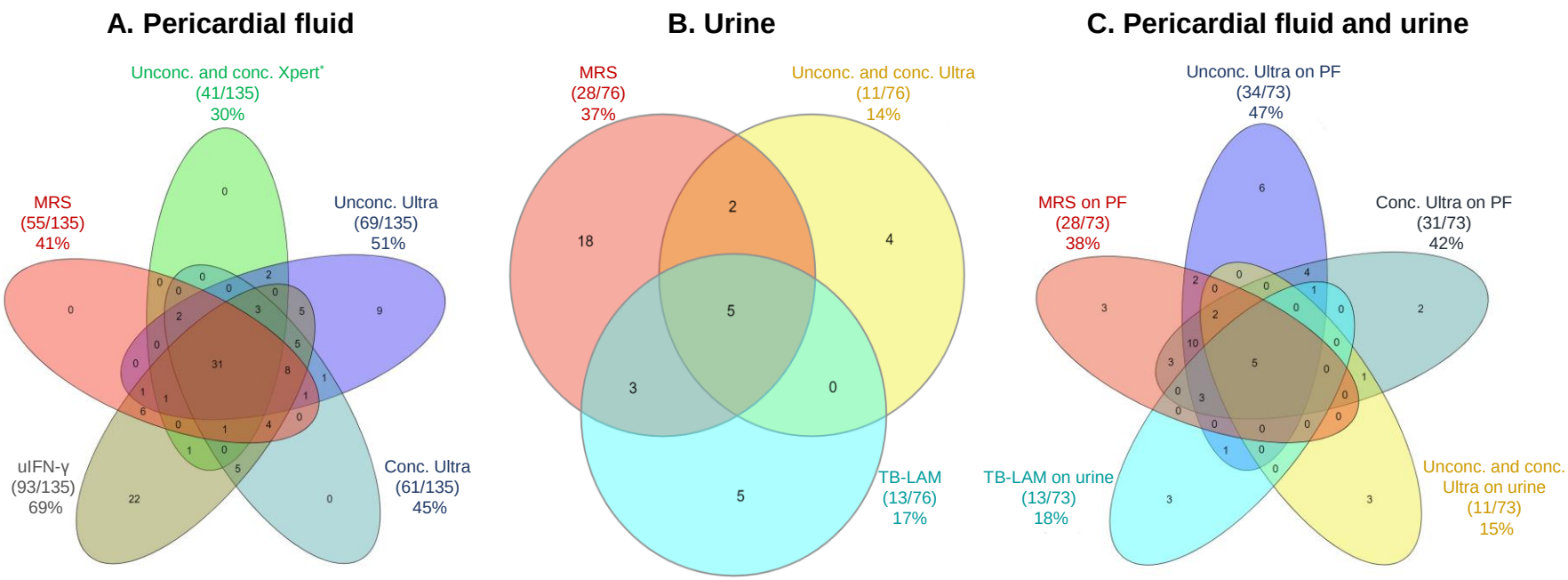
§Current PF or biopsy refers to pericardial fluid or a biopsy collected as part of the study.
¥Previous PF or biopsy refers to pericardial fluid or a biopsy collected within three months of patient recruitment into the study (**Supplement**).
*Unconcentrated Ultra was only done when a concentrated ultra was positive or non-actionable.

Figure 2



§72% (41/57) of PF culture-positives had MTBDR*plus* done (two MTBDR*plus* RIF-resistant, remainder MTBDR*plus* RIF-susceptible). Of the MTBDR*plus* RIF-resistants, 100% (2/2) were detected as resistant by Ultra and 50% (1/2) by Xpert.
¥Of the 25% (15/60) MRS-positives that had culture on biopsy done, 80% (12/15) had MTBDR*plus* (all 12 RIF-susceptible).
*Includes programmatic concentrated Xpert and study unconcentrated Xpert (latter done by the study if former not done programmatically-only PF was used). One MRS-negative had no Xpert testing.
§Both unconcentrated and concentrated study Ultras were only done on PF.
†Study concentrated Ultras were not done in an MRS-positive patient and in two MRS-negatives.

Figure 3



*Includes programmatic concentrated Xpert and study unconcentrated Xpert (latter done by the study if former not done programmatically-only PF was used).

Figure 4

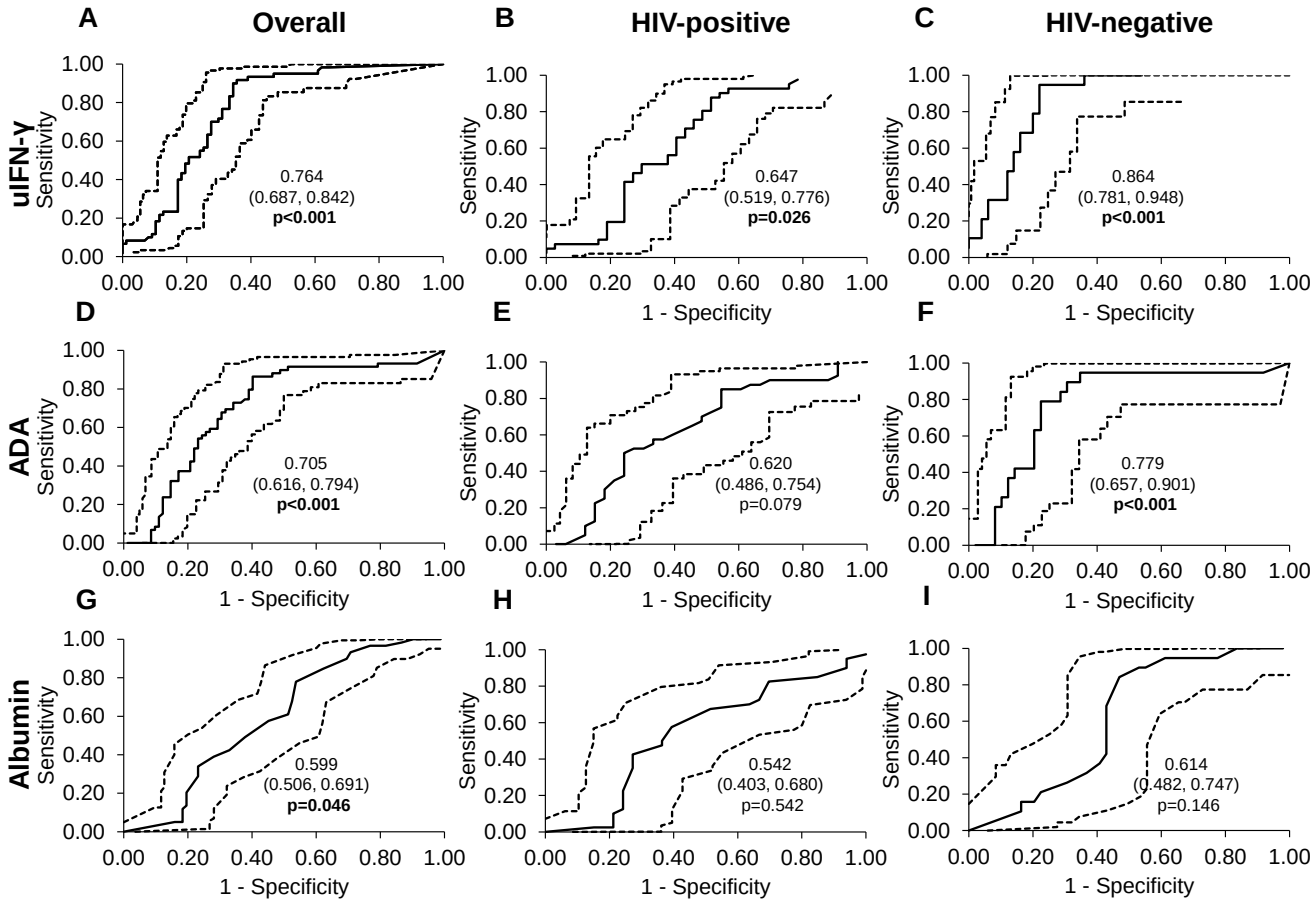


Figure 5

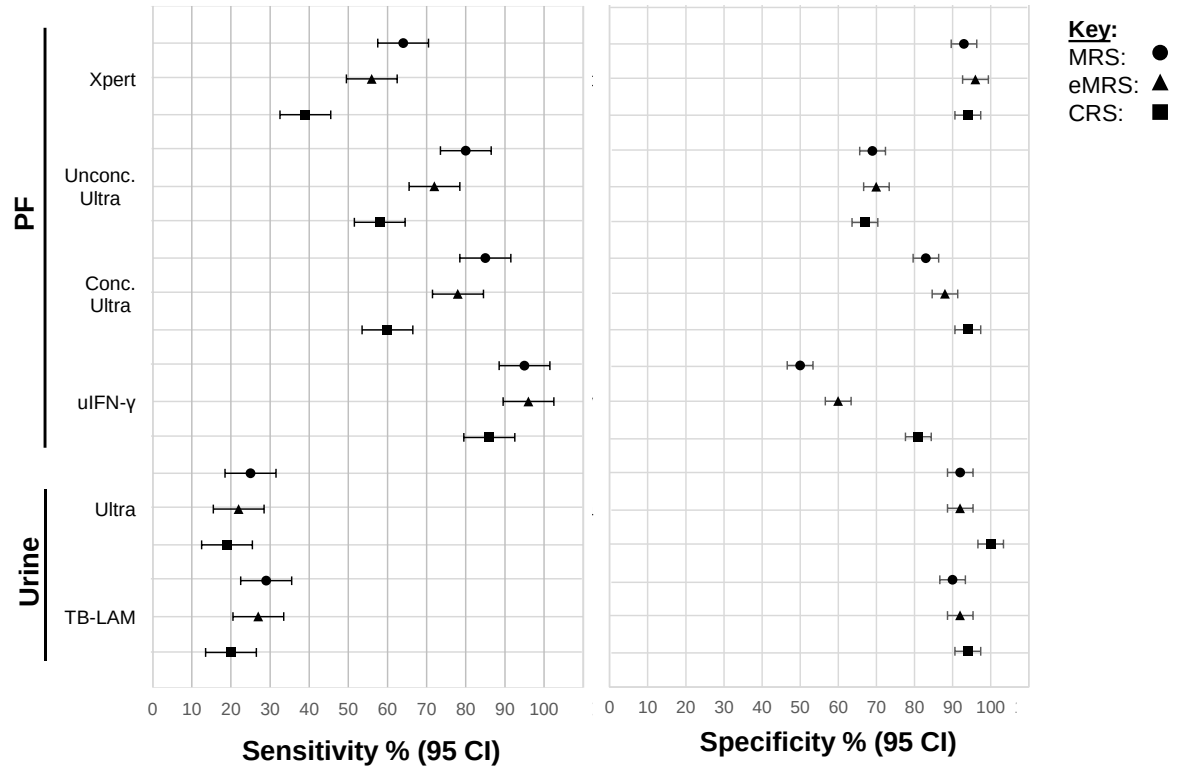


Table 1: Demographic and clinical characteristics by PTB status. Data are n/N (%) or median (IQR).

	Overall (n=150)	**Definite-TB** (n=60)	**Non-TB** (n=90)
Demographics			
Age (years)	41 (34-53)	43 (33-52)	40 (34-54) p=0.986
Female	64/150 (43)	24/60 (40)	40/90 (44) p=0.590
Clinical characteristics			
HIV-positive	79/150 (53)	41/60 (68)	38/90 (42) **p=0.002**
CD4 count (cells/µl)	158 (49-318)	127 (49-323)	183 (70-320) p=0.466
On ART	28/76 (37)	13/41 (32)	15/35 (43) p=0.315
Pericardial tamponade	47/139 (34)	21/53 (40)	26/86 (30) p=0.256
Previous TB	33/147 (22)	16/59 (27)	17/88 (19) p=0.267
Pulmonary TB	32/33 (97)	16/16 (100)	16/17 (94) p=0.340
Extrapulmonary TB	1/33 (3)	0/16 (0)	1/17 (6) p=0.325
Current smoker	31/149 (21)	9/59 (15)	22/90 (24) p=0.177
Symptoms:			
Cough	85/143 (59)	36/55 (65)	49/88 (56) p=0.247
Fever	51/142 (36)	26/56 (46)	25/86 (29) **p=0.035**
Night sweats	71/144 (49)	34/57 (60)	37/87 (43) **p=0.045**
Weight loss	95/145 (66)	41/56 (73)	54/89 (61) p=0.122
Chest X-ray results:			
Normal	8/134 (6)	4/56 (7)	4/78 (5) p=0.627
Abnormal	126/134 (94)	52/56 (93)	74/78 (95) p=0.627

Cardiomegaly	89/134 (66)	38/56 (68)	51/78 (65) p=0.765
Pulmonary infiltrates	33/134 (25)	17/56 (30)	16/78 (21) p=0.192
Hilar lymphadenopathy	10/134 (7)	3/56 (5)	7/78 (9) p=0.432
Miliary pattern	7/134 (5)	4/56 (7)	3/78 (4) p=0.398
Pleural effusion	87/134 (65)	41/56 (73)	46/78 (59) p=0.088
TB treatment			
Treatment at time of recruitment Duration (days) [1 (1-3)]	21/150 (14)	9/60 (15)	12/90 (13) p=0.773
Treatment after 12-week follow-up	110/144 (76)	57/58 (98)	53/86 (62) **p<0.001**
Fluid characteristics			
Total volume aspirated (ml)	700 (500-1000)	645 (500-863)	700 (500-1000) p=0.354
Bloody	67/150 (45)	25/60 (42)	42/90 (47) p=0.546
Chylous	1/150 (1)	0/60 (0)	1/90 (1) p=0.413
Purulent	6/150 (4)	0/60 (0)	6/90 (7) **p=0.041**
Serous	43/150 (29)	18/60 (30)	25/90 (28) p=0.768
Serous-sanguineous	33/150 (22)	17/60 (28)	16/90 (18) p=0.126
Fluid biomarkers			
uIFN-γ (pg/ml)	765.6 (1.40-2351)	1610 (796.10-2595.00)	3.9 (0-1404) **p<0.001**
ADA (U/L)	46.10 (21.45-65.70)	58 (43-74)	33 (12-56) **p<0.001**
Albumin (g/L)	21 (16-26)	2 (16-24)	22 (18-28) **p=0.046**
Total protein (g/L)	57 (49-65.50)	57 (51-64)	57 (48-65.00) p=0.851
LDH (U/L)	660 (440-1483)	869 (548-1642)	623 (327-1229)

			p=0.055
Lymphocyte to neutrophil ratio	1.43 (0.55-5.53)	1.40 (0.77-5.30)	1.60 (0.25-8.80) p=0.528

Five patients MRS-unclassifiable and excluded from this table.

Missing data: CD4, 1; ART, 3; previous TB status, 3; current smoker, 1; cough, 7; fever, 8; night sweats, 6; weight loss, 5; chest X-ray, 16 ; TB treatment, 6; total volume aspirated, 7; pericardial tamponade, 11; lymphocyte to neutrophil ratio, 36; lymphocyte/neutrophil ratio incalculable (divided by zero), 7; ADA, 9; total protein, 9; LDH, 9; albumin, 9; uIFN-γ, 3.

Abbreviations: ADA, Adenosine deaminase; ART, antiretroviral therapy; LDH, lactate dehydrogenase; uIFN-γ, unstimulated interferon-γ.

Table 2: Head-to-head diagnostic accuracy of Xpert, Ultra (unconc. and conc.) and uIFN-γ on PF stratified by HIV status using the MRS. Study unconcentrated Ultra has improved sensitivity in HIV-positive patients and lower specificity overall compared to Xpert (specificity improved with concentration but sensitivity does not). uIFN-γ had higher sensitivity and lower specificity compared to study unconcentrated Ultra. The relative performances of programmatic and study Xpert and study unconcentrated Ultra had similar patterns by HIV status. Accuracy versus the CRS is in **Supplementary Tables 6 and 7**. Data are %, 95% CI, and n/N.

	All patients				HIV-negative				HIV-positive			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
	n=135				**n=62/135 (46)**				**n=73/135 (54)**			
Xpert (programmatic conc. and study unconc. pooled)	64 (50, 76) 35/55	93 (84, 97) 74/80	85 (71, 94) 35/41	79 (69, 86) 74/94	65 (38, 86) 11/17	93 (82, 99) 42/45	79 (49, 95) 11/14	88 (75, 95) 42/48	63 (46, 78) 24/38 p=0.912*	91 (77, 98) 32/35 p=0.748*	89 (71, 98) 24/27 p=0.375*	70 (54, 82) 32/46 **p=0.034***
Unconc. Ultra (study)	80 (67, 90) 44/55 p=0.057‡	69 (57, 79) 55/80 **p<0.001‡**	64 (51, 75) 44/69 **p=0.015‡**	83 (72, 91) 55/66 p=0.468‡	71 (44, 90) 12/17 p=0.714‡	64 (49, 78) 29/45 **p=0.001‡**	43 (24, 63) 12/28 **p=0.028‡**	85 (69, 95) 29/34 p=0.773‡	84 (69, 94) 32/38 **p=0.037‡** p=0.243*	74 (57, 88) 26/35 p=0.057‡ p=0.346*	78 (62, 89) 32/41 p=0.251‡ **p=0.003***	81 (64, 93) 26/32 p=0.245‡ p=0.660*
Conc. Ultra (study)	85 (73, 94) 47/55 p=0.449±	83 (72, 90) 66/80 **p=0.043±**	77 (65, 87) 47/61 p=0.099±	89 (80, 95) 66/74 p=0.313±	82 (57, 96) 14/17 p=0.419±	89 (76, 96) 40/45 **p=0.006±**	74 (49, 91) 14/19 **p=0.037±**	93 (81, 99) 40/43 p=0.270±	87 (72, 96) 33/38 p=744± p=0.663*	74 (57, 88) 26/35 p>0.999± p=0.088*	79 (63, 90) 33/42 p=0.954± p=0.674*	84 (66, 95) 26/31 p=0.784± p=0.211*
uIFN-γ (rule-out cut-off 5.1 pg/ml)	95 (85, 99) 52/55 **p=0.022¥**	50 (39, 61) 40/80 **p=0.016¥**	57 (46, 67) 52/92 p=0.354¥	93 (81, 99) 40/43 p=0.140¥	100 (80, 100) 17/17 **p=0.015¥**	60 (44, 74) 27/45 p=0.664¥	49 (31, 66) 17/35 p=0.651¥	100 (87, 100) 27/27 **p=0.038¥**	92 (79, 98) 35/38 p=0.287¥ p=0.233*	37 (21, 55) 13/35 **p=0.002¥** **p=0.043***	61 (48, 74) 35/57 p=0.081¥ p=0.228*	81 (54, 96) 13/16 p>0.999¥ **p=0.020***

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ±Unconc. Ultra (study) vs. conc. Ultra (study), ¥Unconc. Ultra (study) vs. uIFN-γ in patients of the same HIV status (overall, negative, positive).

Within row p-values: *HIV-negative vs. HIV-positive.

Abbreviations: CI, confidence interval; conc., concentrated; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PF, pericardial fluid; PPV, positive predictive value; uIFN-γ, unstimulated interferon-γ; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Table 3: Head-to-head diagnostic accuracy of urinary Ultra (unconc. or conc.) and TB-LAM stratified by HIV status. Urine has low utility for diagnosing TBP. Ultra was done on concentrated urine and, if non-actionable, on an unconcentrated specimen. Urinary Ultra and TB-LAM had similar accuracy in non-head-to-head analyses (**Supplementary Table 12**). Data are %, 95% CI, and n/N.

	All patients				**HIV-negative**				**HIV-positive**			
	n=76				**n=31/76 (41)**				**n=45/76 (59)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Ultra	25 (11, 45) 7/28	92 (80, 98) 44/48	64 (31, 89) 7/11	68 (55, 79) 44/65	11 (0, 48) 1/9	100 (85, 100) 22/22	100 (3, 100) 1/1	73 (54, 88) 22/30	32 (13, 57) 6/19 p=0.243*	85 (65, 96) 22/26 p=0.055*	60 (26, 88) 6/10 p=0.428*	63 (45, 79) 22/35 p=0.368*
TB-LAM	29 (13, 49) 8/28 p=0.763‡	90 (77, 97) 43/48 p=0.726‡	62 (32, 86) 8/13 p=0.916‡	68 (55, 79) 43/63 p=0.946‡	11 (0, 48) 1/9 p>0.999‡	86 (65, 97) 19/22 p=0.073‡	25 (1, 81) 1/4 p=0.171‡	70 (50, 86) 19/27 p=0.804‡	37 (16, 62) 7/19 p=0.732‡ p=0.159*	92 (75, 99) 24/26 p=0.385‡ p=0.502*	78 (40, 97) 7/9 p=0.405‡ p=0.071*	67 (49, 81) 24/36 p=0.737‡ p=0.755*

Within column p-values: ‡ Unconc. and conc. Ultra vs. TB-LAM.

Within row p-values: *HIV-negative vs. HIV-positive.

Abbreviations: Conc., concentrated; CI, confidence interval; NPV, negative predictive value; PF, pericardial fluid; PPV, positive predictive value; TB-LAM, Determine TB-LAM, Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Chapter 4

Site-of-disease Xpert MTB/RIF Ultra and urine tests for the diagnosis of tuberculosis pleuritis in an HIV-endemic setting

Please note in this chapter, TBP refers to TB pleuritis and PF refers to pleural fluid (this is different in chapter 3)

Supplementary material is attached as Appendix III

Abstract

Background: Tuberculosis pleuritis (TBP) remains the second most common manifestation of extrapulmonary TB (EPTB). Data to support Xpert MTB/RIF Ultra (Ultra) on pleural fluid and urine tests, especially when compared to established biomarkers remains limited.

Methods: Adults programmatically investigated for TBP (n=133) underwent programmatic and study administered 1) Xpert (unconcentrated, concentrated) on pleural fluid (PF) and study administered 2) Ultra and 3) Determine TB-LAM (TB-LAM) on urine and 4) on PF, measurement of unstimulated interferon-γ (IFN-γ). The primary analysis used a microbiological reference standard (MRS).

Results: Unconcentrated PF Ultra (study) had similar sensitivity to Xpert (combined unconcentrated and concentrated results) [80% (95% confidence interval: 44, 97) vs. 50% (19, 81); p=0.160] but had a higher diagnostic yield 78% (51/65) compared to Xpert. When concentrated PF was tested and compared to Xpert, Ultra specificity increased [86% (74, 94) vs. 66% (52, 78); p=0.015], but more non-actionable results occurred [12% (12/97) vs. 1% (1/114); p=0.001]. Ultra diagnostic accuracy did not change or was reduced compared to alternate reference standards. Urine-Ultra and TB-LAM yields were 6% (4/65) and 12% (8/65), respectively, but detected TBP cases missed by culture. uIFN-γ (rule-out cut-point of >221.7 pg/ml) had high sensitivity [100% (83, 100)] and moderate specificity 77% (68, 84).

Conclusions: PF Ultra detects more TBP than Xpert but the decreased specificity requires further investigation. Specimen concentration is useful for improved specificity and uIFN-γ in PF is sensitive for TBP diagnosis. While not sensitive, TB-LAM could reduce the need for thoracentesis for TBP diagnosis.

Background

Tuberculosis pleuritis (TBP) is the second most common manifestation of extrapulmonary TB (EPTB)[84], and can account for up 30% of EPTB cases in a TB endemic setting[85]. TBP is one of the most frequent causes of pleural exudates, and occurs more frequently in people living with HIV (PLHIV)[84].

Diagnosis of TBP remains challenging world-wide as it has a spectrum of presentations[16]. Although generally safe, thoracentesis (hospital procedure to collect pleural fluid) can have complications[86]. In South Africa, patients with TBP are therefore often not diagnosed as most patients only have access to primary care facilities[54]. Additionally, TB culture takes 4-6 weeks and has suboptimal sensitivity in TBP specimens[3].

Xpert MTB/RIF (Xpert; Cepheid, USA) is a semi-automated real-time PCR that rapidly detects *Mycobacterium tuberculosis* complex DNA (MTBC) and rifampicin resistance[2,71,72]. A systematic review and meta-analysis of Xpert on pleural fluid (PF) showed a pooled sensitivity of 51% (95% confidence interval; CI: 43, 60) and specificity of 99% (97, 100) versus culture reference standard[87]. More sensitive tests and associated data for the diagnosis of TBP are thus urgently needed.

Xpert MTB-RIF Ultra (Ultra), Xpert's successor, offers improved sensitivity for TBP and has enhanced limit-of-detection[2]. Data on Ultra in PF remains limited, however. A meta-analysis including four studies in China on PF comparing Ultra and Xpert showed higher pooled Ultra sensitivity [78% (63, 87) vs. 42% (28, 59)] and lower specificity [88% (56, 98) vs. 96% (82, 99)] with culture reference standard[37]. One study in South Africa comparing Xpert and Ultra on PF showed similar sensitivity [38% (25, 51) vs. 29% (16, 41)] and similar specificity [99% (97, 100) vs. 99% (97 to 100)] when a composite reference standard (CRS) (including culture

and pleural biopsy histology) was used[38]. The evaluation of Ultra in South Africa thus remains largely unevaluated.

Given the limitations of microbiological tests, concentrations of unstimulated interferon-γ (uIFN-γ), a cytokine produced by Th1 cells in response to infections, may be useful to diagnose TBP. uIFN-γ has been shown to have high sensitivity and specificity for TBP diagnosis[29,88], and to have higher sensitivity than Ultra on PF when a CRS including culture and histology was used[38]. This study is the first to evaluate uIFN-γ versus Ultra using a culture only microbiological reference standard (MRS) as well as alternate reference standards.

Diagnosing TBP with an easily accessible fluid like urine could mitigate the need for a thoracentesis, while more rapidly diagnosing patients. Determine TB LAM Ag (TB-LAM; Abbott, USA) is a lateral flow assay that detects lipoarabinomannan (LAM; a mycobacterial cell wall component) and has been shown to be useful in patients with EPTB[30,33]. Like TB-LAM, Ultra on urine could improve diagnostic yield for EPTB[32,34], however, Urine-Ultra has not been investigated in patients with presumptive TBP.

We undertook a diagnostic accuracy evaluation of Xpert, Ultra and uIFN-γ on PF and Ultra and TB-LAM on urine in patients with presumptive TBP in South Africa. We hypothesised 1) Ultra would show improved sensitivity compared to Xpert and uIFN-γ in PF, 2) concentrating PF would improve Ultra sensitivity, and 3) TB-LAM and Ultra from non-invasive urine would have suitably high yield to mitigate requirements for thoracentesis for TBP.

Methods and materials

Ethics statement

The study was approved by the Stellenbosch University Human Research Ethics Committee (N16/04/050), Tygerberg General Hospital (TGH), Karl Bremmer Hospital (KBH) and Khayelitsha District Hospital (KDH; all WC_2016RP15_762).

Patient recruitment and sampling

144 inpatients (≥18 years) with an ultrasound-identified pleural effusion undergoing programmatic investigation that required thoracentesis at TGH, KBH and KDH in Cape Town, South Africa, were prospectively consecutively recruited from 2 December 2016-5 April 2023. Programmatically administered chest X-rays (CXRs) were read by a radiologist. PF, blood, and urine were collected by research nurse. Patients currently on TB treatment were excluded.

Pleural fluid collection and programmatic testing algorithm

PF and pleural biopsies, if clinically indicated were collected and sent for programmatic testing at the National Health Laboratory Service (NHLS). Excess fluid was obtained for study testing (**Figure 1**) blinded to programmatic results.

Definitions

Patients were designated as definite- or non-TB. For the microbiological reference standard (MRS), definite TBs were culture-positive on PF or a pleural biopsy, and non-TBs culture-negative on PF and, if done, pleural biopsies. The extended microbiological reference standard (eMRS) included microbiological tests on fluid other than PF and biopsies and the composite reference standard (CRS) included treatment information from follow-ups (**Supplementary Table 1**). Actionable results were those that provided clinically useful information (non-actionable results are defined further in **Supplement**). Unless otherwise stated, we present results versus the MRS and, for comparisons with other tests, head-to-head index tests

diagnostic accuracy results. The final Xpert or Ultra results (i.e., after re-testing if initially non-actionable) are included.

Programmatic testing

Pleural fluid and biopsy testing

Culture: Briefly, 5-7.5 ml PF or biopsy were N-acetyl-l-cysteine–sodium hydroxide (NALC-NaOH)-decontaminated, centrifuged, resuspended in phosphate buffer, and 500 µl inoculated into a MGIT960 tube. Genotype MTBDR*plus* (Hain Lifesciences, Germany) was done on culture isolates for speciation and drug susceptibility testing[76].

Xpert and Ultra: Xpert (version 1; Cepheid, USA) was done programmatically from 25 January 2017–9 April 2018 and replaced with Ultra (version 1) thereafter. Sample reagent buffer (1.5 ml) was added to 500 µl (3:1) PF or biopsy and 2 ml used for Xpert or Ultra[57,58]. The programmatic algorithm for culture and PCR differed if the patient had had a previous PF collected within three months prior (**Supplementary**, **Figure 1**).

Cytochemistry: Adenosine deaminase (ADA) was quantified using the Diazyme Adenosine Deaminase assay (Diazyme laboratories, Inc; USA) and total protein, lactate dehydrogenase (LDH) and albumin were quantified using the SYNCHRON system (Beckman Coulter, USA).

Study testing

Pleural fluid

Xpert and Ultra: Two 700 µl unconcentrated PF aliquots were each freshly tested with Xpert (started by the study when the programme switched to Ultra) and Ultra (both version 3) using 1.4 ml sample reagent buffer (2:1 volume ratio) as per manufacturer's instructions[57,58]. Volume permitting, concentrated Ultra was done after centrifuging 20 ml fresh PF (1711×*g*, 10 min, room temperature) and removing supernatant until 700 µl remained, from which the resuspended pellet was treated with 1.4 ml sample reagent (2:1). Xperts and Ultras with non-actionable results were repeated once volume-permitting.

Unstimulated interferon-γ: IFN-γ concentrations (without antigen stimulation) were measured in duplicate using filter-sterilised supernatant (0.45 µm followed by 0.22 µm filtration; Thermo Scientific) from 1.4 ml of centrifuged (20124×*g*, 15 min, room temperature) PF after one freeze-thaw cycle. A standard curve and background controls were included in each ELISA plate tested using the Human IFN-γ ELISAPRO kit (Mabtech, Sweden) according to manufacturer's instructions[77].

Urine-Ultra and TB-LAM

Freshly collected urine (5-20 ml) was centrifuged (2862×*g*, 15 min, room temperature) and supernatant removed until ~700 µl remained. The pellet was resuspended with 1.4 ml sample reagent buffer (2:1)[58] and tested by Ultra. 700 µl of unconcentrated urine was tested by Ultra when the concentrated Ultra was positive or had a non-actionable result. ~60 µl unconcentrated urine was tested using TB-LAM[78].

Statistical analysis

Proportion tests[59] were done using STATA version 16.0 (StataCorp, USA) and GraphPad Prism version 8.0.1 (GraphPad Software, USA). Euler diagrams were made with InteractiVenn[60]. Differences in diagnostic accuracy metrics were calculated using proportion

tests or McNemar's test. STARD guidelines were followed[61]. Diagnostic yield was calculated as, of people with a positive test result [study Ultra on PF (unconcentrated or concentrated), study Xpert on PF (unconcentrated or concentrated), programmatic PF culture, programmatic culture on pleural biopsy, programmatic smear microscopy, culture, concentrated Ultra and concentrated Xpert-all on separate non-site-of-disease fluid, study Urine-Ultra (concentrated or unconcentrated), study urine TB-LAM], the proportion positive by one specific test.

Patient treatment and follow-up

Treatment decisions were programmatic (no study results reported). Patients were telephonically followed-up at least 12 weeks post-recruitment. TB treatment initiation status was recorded and, if started, patients self-reported treatment response at follow-up. Patients were classified as loss-to-follow-up if at least two calls and messages were unsuccessful.

Results

Patient characteristics

Of 133 patients, 9% (12/133) were MRS-unclassifiable and, of those that could be classified, 15% (18/121) had definite TB and 85% (103/121) were non-TB. Patient characteristics are in **Table 1**.

Xpert and Ultra on PF

An overview of Xpert and Ultra results according to the MRS and whether PF was concentrated is in **Figure 2**. Ultra detected more people than Xpert as positive, including many MRS-negatives.

Non-actionable results: Unconcentrated Xpert had an initial non-actionable result rate of 5% (4/87) [only includes study Xpert results as programmatic Xpert non-actionable results unavailable] and, in those in whom sufficient volume existed, 100% (2/2) were non-actionable after retesting (**Supplementary Table 3**). The non-actionable rate of unconcentrated Xpert and unconcentrated Ultra were similar [5% (4/87) vs. 1% (1/114); p=0.093], with concentrated Ultra increasing to 12% (12/97); p=0.001] (**Supplementary Table 3**). In concentrated Ultras with sufficient volume for a re-test, 100% (1/1) were negative (unconcentrated Ultra had insufficient PF for a re-testing).

Sensitivity and specificity of Xpert and Ultra

Overall: When unconcentrated Ultra was compared to Xpert (**Table 2**), sensitivity was similar [80% (95% confidence interval: 44, 97) vs. 50% (19, 81; p=0.160)] and specificity was lower [66% (52, 78) vs. 98% (90, 100; p<0.001]. Compared to unconcentrated Ultra, concentrated Ultra had similar sensitivity [100% (69, 100) vs. 80% (44, 97); p=0.136] and increased specificity [86% (74, 94) vs. 66% (52, 78); p=0.015]. In the non-head-to-head analysis (**Supplementary Table 4**), unconcentrated Ultra had increased sensitivity compared to Xpert

[79% (49, 95) vs. 38% (14, 68); p=0.034)] and decreased specificity [70% (59, 79) vs. 97% (91, 100); p<0.001)]. Analyses using the eMRS had similar conclusions, however, when the CRS was used, sensitivity loss for both Xpert and Ultra occurred while specificity was similar (**Supplementary Table 5**).

In people without HIV: Unconcentrated Ultra had similar sensitivity [50% (1, 99) vs. 0% (0, 84); p=0.248] and decreased specificity [62% (46, 75) vs. 98% (89, 100); p<0.001] to Xpert (**Table 2**). When concentrated and unconcentrated Ultras were compared, sensitivity was similar [100% (16, 100) vs. 50% (1, 99); p=0.248] and specificity increased [85% (72, 94) vs. 62% (46, 75); p=0.010]. Non-head-to-head analyses had similar conclusions (**Supplementary Table 4**).

In PLHIV: Unconcentrated Ultra had similar sensitivity [88% (47, 100) vs. 63% (24, 91); p=0.248] and similar specificity [89% (52, 100) vs. 100% (9, 9); p=0.304] to Xpert. When concentrated and unconcentrated Ultras were compared, sensitivity [100% (63, 100) vs. 88% (47, 100); p=0.302] and specificity [89% (52, 100) vs. 89% (52, 100); p>0.999] were similar. Non-head-to-head analyses had similar conclusions (**Supplementary Table 4**).

Comparisons of individual tests in people with or without HIV: Sensitivity and specificity were similar in all tests but PPV was increased for all tests in PLHIV (**Table 2**). Non-head-to-head analyses had similar conclusions (**Supplementary Table 4**).

Trace recategorization strategies: When unconcentrated Ultra traces were excluded, sensitivity was similar [+20% (-5, 45); p=0.334] and specificity increased [+25% (10, 41); p=0.006] (**Supplementary Table 6**). When these results were rather reclassified as negative, sensitivity was similar [-30% (-70, 10), p=0.160] and specificity increased [+29% (15, 42), p<0.001]. When concentrated Ultra traces were excluded, sensitivity and specificity were similar, and when reclassified as negative, sensitivity decreased [-40% (-70, -10); p=0.025] and specificity increased [+11% (0, 21); p=0.047].

Ultra inhibition: Sample processing control (SPC) C_T values showed higher inhibition in study concentrated Ultras versus study unconcentrated Ultras [25.70 (IQR: 25.00-26.43) vs. 25.10 (24.28-26.00); p<0.001] (**Supplementary Figure 1A**).

Ultra rpoB C_{Tmin} and IS6110/IS1081 C_T: Study concentrated Ultra had lower *rpo*B C_{Tmin} compared to unconcentrated Ultra [24.70 (IQR: 23.10-26.50) vs. 26.90 (23.00-30.30); p=0.047] and higher IS*6110*/IS*1081* C_T [26.00 (25.30-27.03) vs. 21.80 (19.68-25.18); p=0.007] values (**Supplementary Figure 1B and 1C)**.

Relationship with bacterial load: Xpert, study unconcentrated and study concentrated Ultra *rpo*B C_{Tmin} correlated with culture time-to-positivity (TTP), however, Ultra IS*6110*/IS*1081* C_T only had a positive linear correlation in study concentrated Ultra (and not study unconcentrated Ultra) (**Supplementary Figure 2A-E**).

Rifampicin (RIF) results: Both Xpert and Ultra (unconcentrated and concentrated) correctly identified RIF-susceptibility in all patients that had programmatic drug susceptibility testing (MTBDR*plus*) done (**Supplementary Table 7**).

Programmatic Ultras vs. study Ultras: When programmatic concentrated Ultra and study unconcentrated Ultras were compared, study unconcentrated Ultras had fewer positive results [-15% (95% confidence interval: -31,1); p=0.046] (**Supplementary Table 8**). When programmatic concentrated Ultras were compared to study concentrated Ultras and study concentrated Ultras were compared to unconcentrated Ultras, positivity rates were similar.

<u>uIFN-γ on PF</u>

uIFN-γ sensitivity and specificity

Optimal cut-points: Of the 133 patients enrolled, 15% (20/133) did not have uIFN-γ done. Optimal rule-in, rule-out and Youden's index cut-point concentrations for uIFN-γ on PF to maximise diagnostic accuracy were determined using the ROC curve shown in **Figure 4A**. The

area under the curve (AUC) for all patients was 0.925. We prioritised sensitivity for diagnostic accuracy analyses (**Table 2**) by using a cut-point of >221.7 pg/ml, which corresponds to a sensitivity of 100% (83, 100) and 77% (68, 84) specificity. Specificity being prioritised and Youden's index can be seen in the supplement. The AUC and cut-points for PLHIV and people without HIV (**Figure 4B** and **4C)** can be seen in supplementary results (**Supplementary Table 2**).

Diagnostic accuracy (head-to-head): When unconcentrated Ultra and uIFN-γ were compared, uIFN-γ (221.7 pg/ml) had similar sensitivity [100% (69, 100) vs. 80% (44, 97); p=0.136] and specificity [77% (64, 87) vs. 66% (52, 78); p=0.210] (**Table 2**). When unconcentrated Ultra and uIFN-γ were compared in PLHIV, uIFN-γ had similar sensitivity [100% (63, 100) vs. 88% (47, 100); p=0.302] and specificity [67% (30, 93) vs. 89% (52, 100); p=0.257]. In HIV-negative patients, uIFN-γ had similar sensitivity [100% (16, 100) vs. 50% (1, 99); p=0.248] and specificity [79% (64, 89) vs. 62% (46, 75); p=0.071].

When PLHIV were compared to patients without HIV, PLHIV had similar sensitivity, specificity and NPV, while PPV increased in PLHIV [73% (39, 94) vs. 17% (2, 48); p=0.007]. When actionable results were compared for: 1) MRS, 2) Xpert, 3) unconcentrated Ultra, 4) concentrated Ultra and 5) uIFN-γ **(**rule-out cut-off 221.7 pg/ml**)**, 15% (10/66), 9% (6/66), 41% (27/66), 27% (18/66), 35% (23/66) were positive by each test (**Figure 3A**); unconcentrated Ultra followed by uIFN-γ had the highest yield.

Urine results

Tests on urine detected TBP in some cases missed by tests on PF

Of concentrated urine specimens tested with Ultra (n=34), 12% (4/34) were non-actionable and 100% (4/4) resolved to actionable when unconcentrated urine was tested (25% (1/4) unconcentrated urines became Ultra-positive). When Urine-Ultra was compared to TB-LAM using the MRS (*Overall, head-to-head*; n=31) (**Table 3**), Ultra had similar sensitivity [50%

(95% confidence interval: 7, 93) vs. 75% (19, 99); p=0.465] and similar specificity [96% (81, 100) vs. 81% (62, 94); p=0.083]. Data was unchanged after stratification by HIV status. Ultra detected 3% (1/31) TBP not detected by the MRS (**Figure 3B**), and TB-LAM detected 16% (5/31) TBP cases missed by the MRS (**Figure 3B**).

When actionable results were compared for the 1) MRS on PF, 2) unconcentrated and concentrated Ultra on urine, 3) TB-LAM on urine, 13% (4/31), 10% (3/31), 26% (8/31) were positive by each test (**Figure 3B**; TB-LAM had the highest yield**)**.

When actionable results were compared for the 1) MRS on PF, 2) unconcentrated Ultra on PF, 3) concentrated Ultra on PF, 4) unconcentrated and concentrated Ultra on urine and 5) TB-LAM on urine, 19% (3/16), 31% (5/16), 25% (4/16), 13% (2/16), 19% (3/16) were positive by each test (**Figure 3C**; unconcentrated Ultra on PF followed by concentrated Ultra on PF had the highest yield).

Overall (patients who did or did not have the test attempted)

Study Ultra (unconc. and conc.) on PF had the highest yield 78% (51/65) followed by culture on PF 28% (18/65), Xpert (unconc. and conc.) on PF 12% (8/65), TB-LAM on urine 12% (8/65) and programmatic Ultra on non-site disease fluid 12% (8/65) (**Supplementary Table 11**). Culture on non-site-of-disease fluid followed by Ultra on urine, smear microscopy and programmatic Xpert on non-site-disease fluid had low yields 9% (6/65), 6% (4/65), 5% (3/65) and 2% (1/65) respectively.

Overall (in patients who had the test attempted)

Study Ultra (unconc. and conc.) on PF had the highest yield 45% (51/113), followed by TB-LAM on urine 25% (8/32) and culture on PF 15% (18/121) had the highest diagnostic yield (**Supplementary Table 11**). Ultra on urine followed by Xpert on PF, programmatic Ultra and culture, programmatic Ultra, culture, smear microscopy and programmatic Xpert on non-site-

of-disease fluid pleural biopsy had low yields of 11% (4/35), 6% (8/133), 5% (6/133) and 2% (3/133) and 1% (1/133) and 0% (0/3) respectively.

Comparison of unconcentrated Ultra true-positives and false-positives per the MRS

True-positives were more likely to be HIV-positive than false-positives [82% (95% interquartile range: 9/11) vs. 11% (3/27), p<0.001] (**Supplementary Table 9**). More true-positives showed a miliary pattern on their chest X-ray [20% (2/10) vs. 0% (0/24), p=0.024] than false-positives and were Xpert-positive [50% (5/10) vs. 9% (2/23), p=0.008], eMRS-positive [100% (11/11) vs. 0% (0/27), p<0.001] and CRS-positive [100% (11/11) vs. 37% (10/27), p<0.001]. FPs were more likely to show a pleural effusion on chest X-rays [100% (24/24) vs. 80% (8/10), p=0.024], had higher IS*6110*/IS*1081* C_T [27.00 (Interquartile range: 24.00-27.90) vs. 23.00 (19.50-26.40), p=0.011] and lower uIFN-γ concentrations [0 pg/ml (Interquartile range:0-435) vs. 4432 (2359-6129); p<0.001]. The characteristics of false-positives per patient information is in **Supplementary Table 10**.

Diagnostic Yield

Overall (patients who did or did not have the test attempted)

Study Ultra (unconc. and conc.) on PF had the highest yield 78% (51/65) followed by culture on PF 28% (18/65) and Xpert (unconc. and conc.) on PF 12% (8/65) (**Supplementary Table 11**). TB-LAM had a yield of 12% (8/65) and urine-Ultra had a yield of 6% (4/65). Tests on non-site-of-disease fluid and culture on biopsy had low yields respectively. Xpert (unconc. and conc.) and culture on PF had higher yields in people without HIV compared to PLHIV [30% (6/20) vs. 4% (2/45); p=0.004] and [70% (14/20) vs. 9% (4/45); p<0.001] respectively. Smear microscopy, culture, programmatic Ultra and programmatic Xpert on non-site-of-disease fluid had higher yields in people without HIV.

Overall (in patients who had the test attempted)

Smear microscopy on non-site-of-disease fluid had the highest yield 100% (3/3), followed by study Ultra (unconc. and conc.) on PF 45% (51/113), programmatic Xpert 33% (1/3) and programmatic Ultra 25% (8/32) (both on non-site-of-disease fluid) and TB-LAM on urine 25% (8/32) (**Supplementary Table 11**).

Patient treatment status at follow-up

96% (128/133) of patients were followed-up [median (Interquartile range: 21 (14-63) weeks since recruitment] and 35% (45/127) of those had initiated TB treatment. Of these, 31% (14/45) were classified as definite TB and 58% (26/45) as non-TB per the MRS [11% (5/45) were unclassifiable]. Of the definite TBs, 73% (8/11) were unconcentrated Ultra-positive and for non-TBs, 40% (10/25) were unconcentrated Ultra-positive. Regarding the clinical status in patients who started treatment, 96% (43/45) reported treatment completion and, of these, 86% (37/43) reported feeling clinically well. 31% (40/128) patients were documented to have died, of whom 20% (8/40) who were not placed on treatment were exclusively unconcentrated Ultra-positive.

Discussion

Our key findings are: 1) Ultra on unconcentrated PF, compared to Xpert, had similar sensitivity but higher diagnostic yield overall (tests were unaffected by HIV and the diagnostic accuracy compared to eMRS or CRS was similar); 2) Ultra on unconcentrated PF showed decreased specificity compared to Xpert, but increased specificity when PF was concentrated, but concentrated Ultra came with an increase in non-actionable results; 3) Exclusion of trace Ultra results did not change sensitivity but increased specificity, while reclassifying trace results as negative decreased sensitivity and increased specificity; 4) uIFN-γ has high sensitivity on PF but has moderate specificity; 5) Urine tests detected few TBP cases but did detect 6% TBP that tests on PF did not, and thus could reduce the proportion of patients who require thoracentesis for TBP diagnosis. These data demonstrate the increased yield of Ultra for TBP diagnosis with the inclusion of trace-positive results.

Unconcentrated Ultra on PF had similar sensitivity but had a higher diagnostic yield compared to Xpert, suggesting that Ultra can be used as a rapid initial test for TBP diagnosis. Additionally, when unconcentrated Ultra was compared to Xpert, an overall loss in specificity was observed (this was unchanged when the eMRS and CRS was used). This was different to a systematic review and meta-analysis of Ultra compared to Xpert which showed that when culture was used as the reference, Ultra had improved sensitivity and lower specificity compared to Xpert[89]. However, all these studies took place in China that has lower TB/HIV prevalence. Interestingly, when a study in South Africa compared Ultra to Xpert using a CRS (including culture and histology), Ultra sensitivity was similar to that of Xpert, consistent with our findings[38].

Loss in Ultra specificity could be due to the flawed nature of reference standards in non-sputum specimens[24] and was not driven by previous TB as seen in pulmonary specimens[53]. Moreover,

when PF was concentrated, improved specificity was observed compared to unconcentrated Ultra. This was also observed in pericardial fluid (chapter 3). We speculate that concentrated Ultra's improved specificity might be due to increased PCR inhibitors (high SPC C_T) changing concentrated Ultra false-positives (that has a low bacterial count) into true-negatives, thereby increasing specificity. Concentration of PF does come with an increase in non-actionable concentrated results compared to unconcentrated Ultra, which could be due to the concentration of inhibitory agents in PF. The benefit of PF concentration might thus be negated by increased non-actionable results which is noteworthy as programmatic Ultras in South Africa uses concentrated PF.

Critically, when trace results were excluded as an analytical approach, specificity increased, and sensitivity remained similar. However, if trace results were reclassified as negative, specificity improvement came with a loss in sensitivity, suggesting that trace exclusion is the superior strategy for handling trace results if used clinically or in research.

uIFN-γ has high sensitivity in diagnosing TBP when the rule-out cut-off of 221.7 pg/ml was used. This is similar to a systematic review that showed high sensitivity when evaluating unstimulated interferon gamma on PF[37]. High specificity was observed in this systematic review and meta-analysis compared to the moderate specificity seen in our study, which could be due the use of different unstimulated IFN-γ assays and reference standards being used. This data suggests that laboratories with sufficient financial capacity and instrumentation available could use uIFN-γ for TBP diagnoses.

The use of sensitive tests on non-invasive specimens for TBP diagnosis are still needed. The sensitivity of urine-Ultra and TB-LAM were not significantly different in this cohort, and only detected TB in PLHIV. This suggests that for TB pleuritis diagnosis, particularly in resource limited settings with a high HIV prevalence, TB-LAM should be used as an initial test followed

by more invasive testing. Nonetheless TB-LAM and Urine-Ultra identified TB missed by culture, suggesting universal urine testing could reduce the number of patients undergoing thoracentesis in a subset of patients.

Our data represent one of the first studies evaluating molecular, microbiological and biomarker tests for TBP diagnosis in ahead-to-head analysis. Pragmatic limitations of our study include that programmatic Xpert (which was concentrated) was not always done, therefore study Xpert (which was unconcentrated) was done and a combination of the two was used for Xpert analyses. PF is decontaminated at the NHLS (where programmatic tests are done) and further diluted with more sample reagent (SR) buffer for Xpert and Ultra than recommended[57,58], which may underestimate their sensitivity.

In conclusion, with the inclusion of trace results, Ultra confirms TBP in more people with than Xpert, but this comes with a loss in specificity. Concentrating PF improves Ultra's specificity, but the increased non-actionable result rate may negate this benefit. uIFN-γ has high sensitivity and moderate specificity but does not provide significant diagnostic benefit over Ultra on PF. TB-LAM should be used as initial test for TBP diagnosis followed by tests on PF if negative, which could reduce the need for invasive sampling in a subset of patients. Our data support the use of Ultra for TBP diagnosis, while emphasising the urgent need for more sensitive tests on non-invasively collected specimens such as urine.

Acknowledgments

The authors thank the study participants, National Health Laboratory Services, Tygerberg Hospital, Dawn Davis, Ruth Wilson, Nandi Niemand, Roxanne Higgit, Selisha Naidoo, Dr Megan Hendrikse and Zaida Palmer. Research reported in this publication was supported by the South African Medical Research Council. GT acknowledges funding from the EDCTP2 programme supported by the European Union (RIA2018D-2509, PreFIT;RIA2018D-2493, SeroSelectTB; RIA2020I-3305, CAGE-TB) and the National Institutes of Health (D43TW010350; U01AI152087; U54EB027049; R01AI136894). CN is supported by the Fogarty International Center of the National Institutes of Health (K43TW012302). The degree from which this study emanated was made possible through funding by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the SAMRC Internship Scholarship Programme. We acknowledge the support of the South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. The content is solely the responsibility of the authors and does not necessarily represent the official views of any funders. Cepheid donated cartridges but had no role in study design or analysis.

Figure legends

Figure 1: Specimen collection and diagnostic procedures done in participants with suspected TB pleuritis. Abbreviations: conc., concentrated; PF, pleural fluid; TB, tuberculosis; TB-LAM, Determine TB-LAM; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF; uIFN-γ, unstimulated interferon-γ.

Figure 2: Summary of test results done programmatically (culture, concentrated Xpert, concentrated Ultra) and by the study (unconcentrated and concentrated Ultra, unconcentrated Xpert when the programme did not do Xpert) are shown, from either PF or biopsies. Ultra detected more TBP in culture-positive and culture-negative specimens compared to Xpert. Concentration increased Ultra positivity in MRS-negatives. Data are n/N (%). Abbreviations: conc., concentrated; PF, pleural fluid; RIF, rifampicin; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

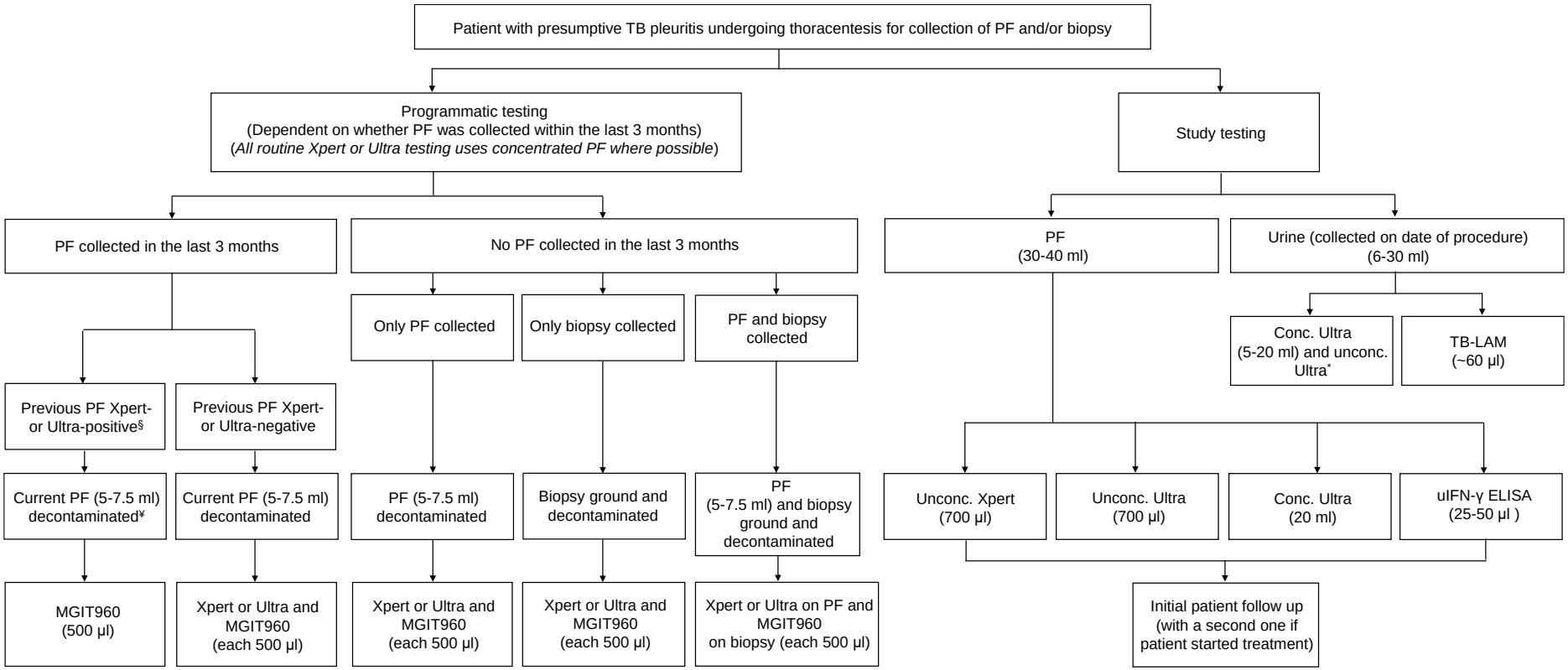
Figure 3: Euler diagrams showing positive results from tests done on PF and/or urine versus the MRS. **(A)** Xpert, study Ultra (unconcentrated and concentrated) and uIFN-γ (cut-off 221.7 pg/ml) on PF. Both PF study unconcentrated Ultra and uIFN-γ were positive in patients not detected by Xpert, study concentrated Ultra, and the MRS. **(B)** MRS, urine study Ultra (unconcentrated and concentrated) and urine TB-LAM results irrespective of HIV status. Urinary TB-LAM detected patients missed by the MRS and urine on Ultra. **(C)** MRS, study unconcentrated Ultra on PF, study concentrated Ultra on PF and, on urine, study unconcentrated and concentrated Ultra and TB-LAM (irrespective of HIV status). PF study unconcentrated Ultra followed by study concentrated Ultra on PF detected patients missed by the MRS and urine-based tests. Urine tests detected TBP in some cases missed by PF tests. Data are n/N (%). Abbreviations: conc., concentrated; MRS, microbiological reference

standard; PF, pleural fluid; TB-LAM, Determine TB-LAM; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF; uIFN-γ, unstimulated interferon-γ.

Figure 4: Receiver operator characteristic (ROC) curve analysis of uIFN-γ on pleural fluid in patients with suspected TB. Values are AUCs and 95% confidence intervals. **(A)** AUC of uIFN-γ in all patients, **(B)** in HIV-positive patients and **(C)** HIV-negative patients. uIFN-γ on pleural fluid has good diagnostic accuracy. Abbreviations: AUC, area under the receiver operating characteristic curve; ROC, receiver operator characteristics; TB, tuberculosis; uIFN-γ, unstimulated interferon gamma. Abbreviations: AUC, area under the receiver operating characteristic curve; CIs, confidence intervals; ROC, receiver operator characteristics; uIFN-γ, unstimulated interferon-γ.

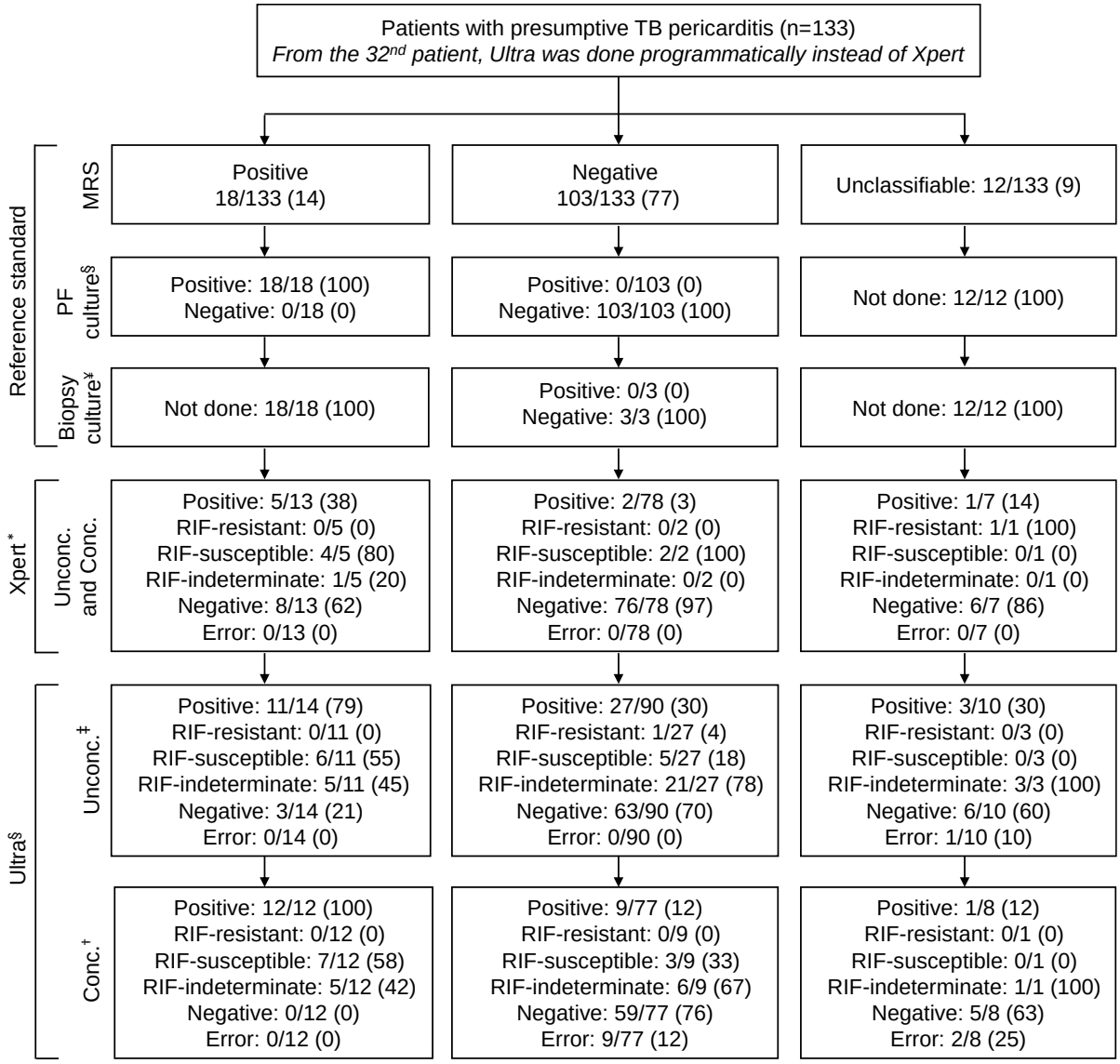
Figure 5: Sensitivity and specificity of tests on PF and urine compared to the MRS, eMRS and CRS. uIFN-γ and Ultra on PF had high sensitivity, and when compared to Xpert, Ultra had lower specificity. Tests on urine had moderate sensitivity. Abbreviations: CI, confidence interval; CRS, composite reference standard; conc., concentrated; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; PF, pleural fluid; uIFN-γ, unstimulated interferon-γ, Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Figure 1



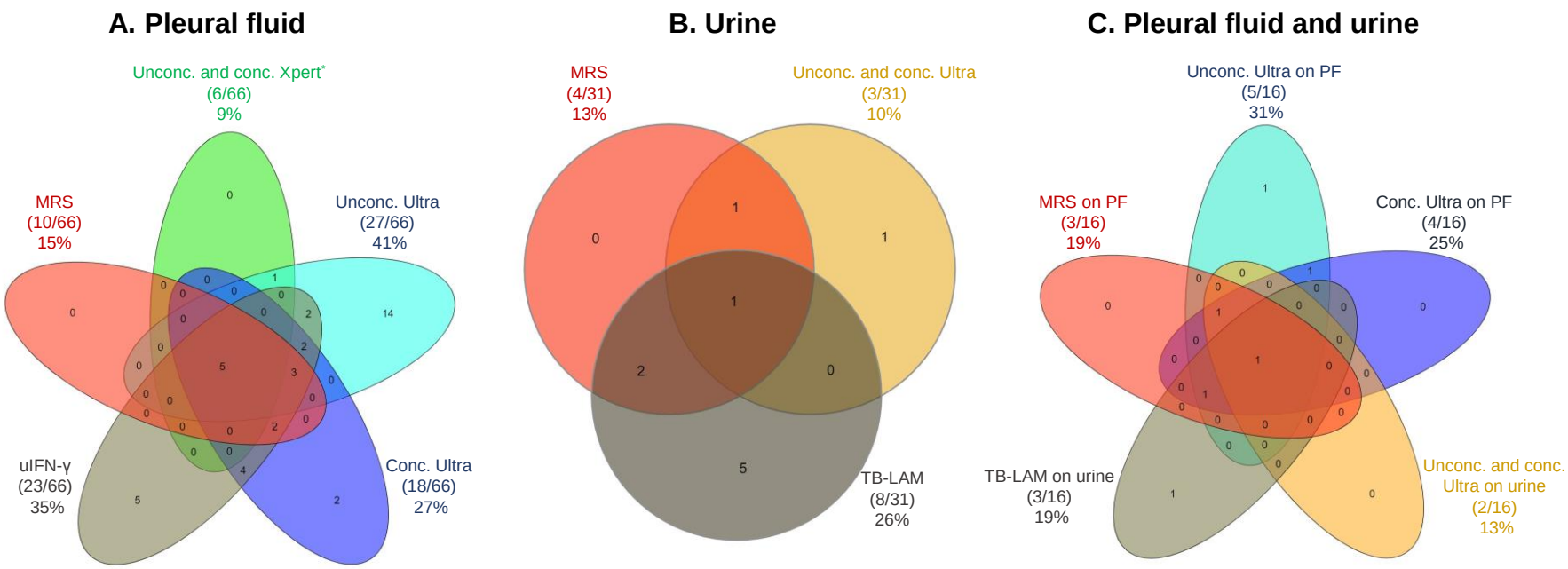
§Current PF or biopsy refers to pleural fluid or a biopsy collected as part of the study.
¥Previous PF or biopsy refers to pleural fluid or a biopsy collected within three months of patient recruitment into the study (**Supplement**).
*Unconcentrated Ultra was only done when a concentrated ultra was positive or non-actionable.

Figure 2



§61% (11/18) of PF culture-positives had MTBDR*plus* done (all MTBDR*plus* RIF-susceptible).
¥97% (100/103) MRS-negatives had no biopsy done.
*Includes routine concentrated Xpert and study unconcentrated Xpert (latter done by the study if former not done programmatically-only PF was used). Five MRS-positives, 25 MRS-negatives and five MRS-unclassifiable patients had no Xpert testing.
§Both unconcentrated and concentrated study Ultras were only done on PF.
‡Four MRS-positives, 13 MRS-negatives and two MRS-unclassifiable patients did not have unconcentrated Ultras done.
†Study concentrated Ultras were not done in six MRS-positives, 26 MRS-negatives and four MRS-unclassifiable patients.

Figure 3



*Includes programmatic concentrated Xpert and study unconcentrated Xpert (latter done by the study if former not done programmatically-only PF was used).

Figure 4

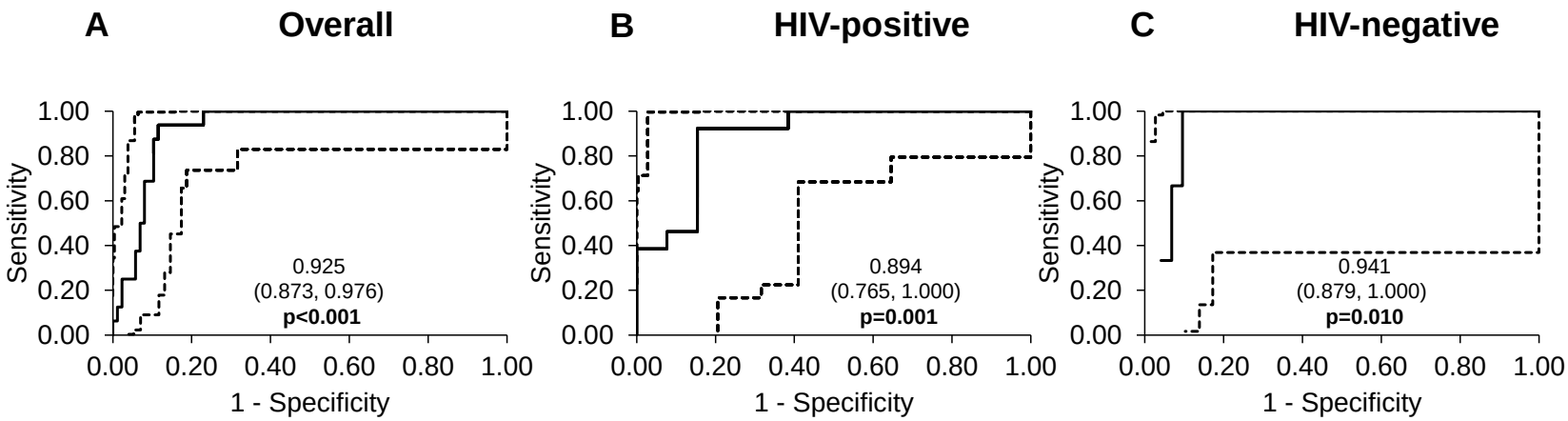


Figure 5

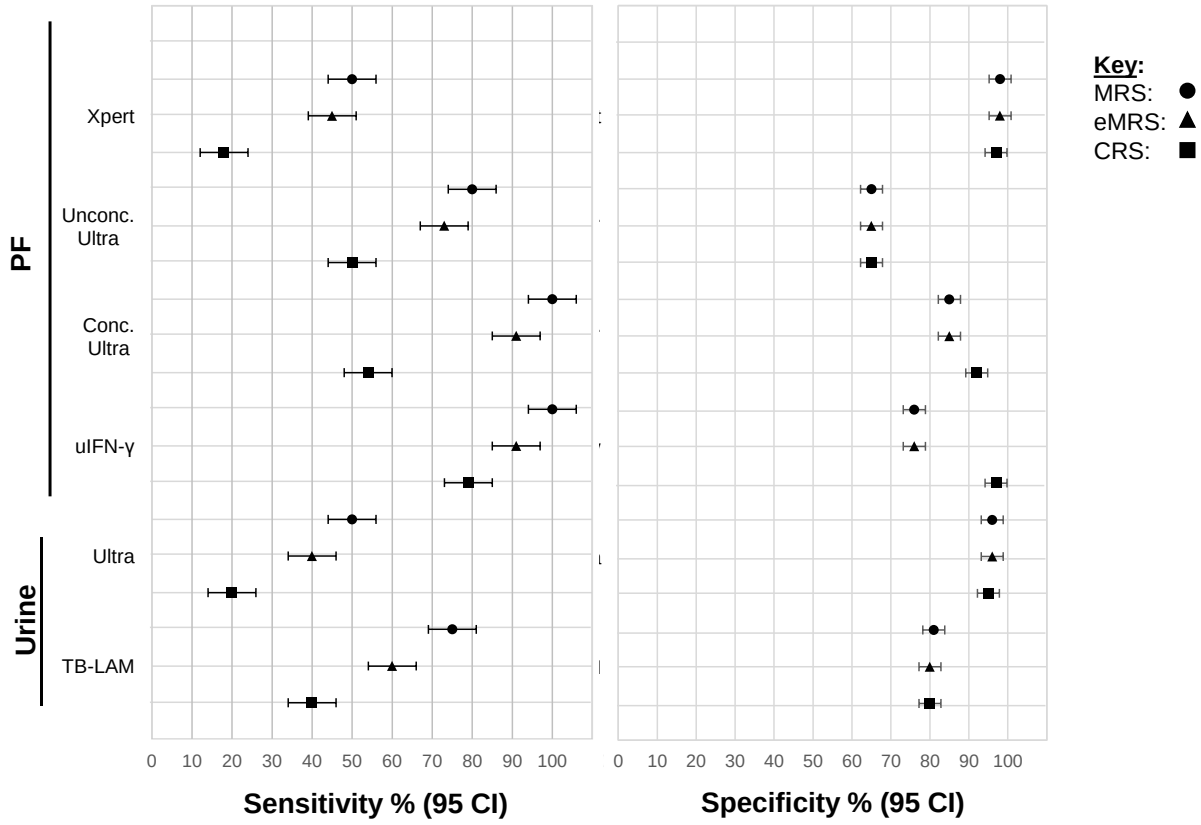


Table 1: Demographic and clinical characteristics by PTB status. Definite-TBs were older, more likely to have HIV, show a miliary pattern on chest X-rays, to have started TB treatment after their 12 week follow-up and have higher uIFN-γ, ADA, protein and LDH concentrations compared to non-TBs. Non-TBs were more likely to have started ART. Data are n/N (%) or median (IQR).

	Overall (n=121)	**Definite-TB** (n=18)	**Non-TB** (n=103)
Demographics			
Age (years)	42 (31-54)	41 (31-53)	56 (41-69) **p=0.002**
Female	61/121 (50)	10/18 (56)	51/103 (50) p=0.636
Clinical characteristics			
HIV-positive	33/120 (28)	14/18 (78)	19/102 (19) **p<0.001**
CD4 count (cells/µl)	166 (89-311)	117 (36-228)	244 (100-430) p=0.131
On ART	14/34 (41)	3/13 (23)	11/19 (58) **p<0.001**
Previous TB	19/120 (16)	4/17 (24)	15/103 (15) p=0.348
Pulmonary TB	18/19 (95)	4/4 (100)	14/15 (93) p=0.596
Extrapulmonary TB	1/19 (5)	0/4 (0)	1/15 (7) p=0.596
Current smoker	25/121 (21)	4/18 (22)	21/103 (20) p=0.859
Symptoms:			
Cough	70/119 (59)	13/17 (76)	57/102 (56) p=0.110
Fever	21/119 (18)	5/17 (29)	16/102 (16) p=0.169
Night sweats	38/118 (32)	8/17 (47)	30/101 (30) p=0.157
Weight loss	84/119 (71)	14/17 (82)	70/102 (69) p=0.250
Chest X-ray results:			
Normal	0/110 (0)	0/15 (0)	0/95 (0) p>0.999

Cardiomegaly	4/110 (4)	0/15 (0)	4/95 (4) p=0.418
Pulmonary infiltrates	19/110 (17)	1/15 (7)	18/95 (19) p=0.242
Hilar lymphadenopathy	0/110 (0)	0/15 (0)	0/95 (0) p>0.999
Miliary pattern	2/110 (2)	2/15 (13)	0/95 (0) **p<0.001**
Pleural effusion	100/110 (91)	14/15 (93)	86/95 (91) p=0.725
TB treatment			
Treatment after 12-week follow-up	40/115 (35)	14/18 (78)	26/97 (27) **p<0.001**
Fluid biomarkers			
uIFN-γ (pg/ml)	5.4 (0.0-1005)	2539 (1556-5396)	3.1 (0-166) **p<0.001**
ADA (U/L)	19.15 (7.8-47.15)	52 (20-64)	17 (7.8-42) **p=0.040**
Total protein (g/L)	45.50 (32.25-55.)	59 (40-63)	44 (31-52) **p=0.028**
LDH (U/L)	370 (182-912)	712 (352-1034)	330 (157-933) **p=0.050**
Lymphocyte to neutrophil ratio	4.10 (0.45-14.59)	3.9 (0.30-7.00)	4.7 (0.46-19.00) p=0.383

Twelve patients were unclassifiable per the MRS.

Missing data: HIV,1; CD4, 6; ART, 1; previous TB status, 1; cough, 2; fever, 2; night sweats, 3; weight loss, 2; chest X-ray, 11 ; TB treatment, 6; uIFN-γ, 18; ADA, 18; total protein, 18; LDH, 18; uIFN-γ, 18; lymphocyte to neutrophil ratio, 53; lymphocyte/neutrophil ratio incalculable (divided by zero); 4.

Abbreviations: ADA, Adenosine deaminase; ART, antiretroviral therapy; LDH, lactate dehydrogenase; uIFN-γ, unstimulated interferon-γ.

Table 2: Head-to-head diagnostic accuracy of Xpert, Ultra (unconc. and conc.) and uIFN-γ on PF stratified by HIV status using the MRS. Study unconcentrated Ultra has similar sensitivity and lower specificity overall compared to Xpert (specificity improved with concentration but sensitivity does not). uIFN-γ had similar sensitivity and specificity compared to study unconcentrated Ultra. The relative performances of all tests had similar patterns by HIV status. Accuracy versus the CRS is in **Supplementary Tables 6**. Data are %, 95% CI, and n/N.

	All patients				HIV-negative				HIV-positive			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
	n=66				**n=49/66 (74)**				**n=17/66 (26)**			
Xpert (programmatic conc. and study unconc. pooled)	50 (19, 81) 5/10	98 (90, 100) 55/56	83 (36, 100) 5/6	92 (82, 97) 55/60	0 (0, 84) 0/2	98 (89, 100) 46/47	0 (0, 97) 0/1	96 (86, 99) 46/48	63 (24, 91) 5/8 p=0.114*	100 (66, 100) 9/9 p=0.659*	100 (48, 100) 5/5 **p=0.014***	75 (43, 95) 9/12 **p=0.020***
Unconc. Ultra (study)	80 (44, 97) 8/10 p=0.160‡	66 (52, 78) 37/56 **p<0.001‡**	30 (14, 50) 8/27 **p=0.015‡**	95 (83, 99) 37/39 p=0.543‡	50 (1, 99) 1/2 p=0.248‡	62 (46, 75) 29/47 **p<0.001‡**	5 (0, 26) 1/19 p=0.814‡	97 (83, 100) 29/30 p=0.852‡	88 (47, 100) 7/8 p=0.248‡ p=0.236*	89 (52, 100) 8/9 p=0.304‡ p=0.346*	88 (47, 100) 7/8 p=0.411‡ **p<0.001***	89 (52, 100) 8/9 p=0.423‡ p=0.354*
Conc. Ultra (study)	100 (69, 100) 10/10 p=0.136±	86 (74, 94) 48/56 **p=0.015±**	56 (31, 78) 10/18 p=0.082±	100 (93, 100) 48/48 p=0.112±	100 (16, 100) 2/2 p=0.248±	85 (72, 94) 40/47 **p=0.010±**	22 (3, 60) 2/9 p=0.175±	100 (91, 100) 40/40 p=0.245±	100 (63, 100) 8/8 p=302± p>0.999*	89 (52, 100) 8/9 p>0.999± p=0.766*	89 (52, 100) 8/9 p=0.923± **p=0.004***	100 (63, 100) 8/8 p=0.331± p>0.999*
uIFN-γ (rule-out cut-off 221.7 pg/ml)	100 (69, 100) 10/10 p=0.136¥	77 (64, 87) 43/56 p=0.210¥	43 (23, 66) 10/23 p=0.309¥	100 (92, 100) 43/43 p=0.133¥	100 (16, 100) 2/2 p=0.248¥	79 (64, 89) 37/47 p=0.071¥	17 (2, 48) 2/12 p=0.296¥	100 (91, 100) 37/37 p=0.263¥	100 (63, 100) 8/8 p=0.302¥ p>0.999*	67 (30, 93) 6/9 p=0.257¥ p=0.433*	73 (39, 94) 8/11 p=0.436¥ **p=0.007***	100 (54, 100) 6/6 p=0.398¥ p>0.999*

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ±Unconc. Ultra (study) vs. conc. Ultra (study), ¥Unconc. Ultra (study) vs. uIFN-γ in patients of the same HIV status (overall, negative, positive).

Within row p-values: *HIV-negative vs. HIV-positive.

Abbreviations: CI, confidence interval; conc., concentrated; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PF, pleural fluid; PPV, positive predictive value; uIFN-γ, unstimulated interferon-γ; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Table 3: Head-to-head diagnostic accuracy of urinary Ultra (unconc. or conc.) and TB-LAM stratified by HIV status. Urine has low utility for diagnosing TBP. Ultra was done on concentrated urine and, if non-actionable, on an unconcentrated specimen. Data are %, 95% CI, and n/N.

	All patients				**HIV-negative**				**HIV-positive**			
	n=31				**n=20/31 (65)**				**n=11/31 (35)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Ultra	50 (7, 93) 2/4	96 (81, 100) 26/27	67 (9, 99) 2/3	93 (76, 99) 26/28	0	95 (75, 100) 19/20	0 (0, 97) 0/1	100 (82, 100) 19/19	50 (7, 93) 2/4	100 (59, 100) 7/7 p=0.547*	100 (16, 100) 2/2 p=0.083*	78 (40, 97) 7/9 **p=0.033***
TB-LAM	75 (19, 99) 3/4 p=0.465‡	81 (62, 94) 22/27 p=0.083‡	38 (9, 76) 3/8 p=0.387‡	96 (78, 100) 22/23 p=0.673‡	0	75 (51, 91) 15/20 p=0.077‡	0 (0, 52) 0/5 p>0.999‡	100 (78, 100) 15/15 p>0.999‡	75 (19, 99) 3/4 p=0.465‡	100 (59, 100) 7/7 p>0.999‡ p=0.143*	100 (29, 100) 3/3 p>0.999‡ **p=0.005***	88 (47, 100) 7/8 p=0.600‡ p=0.162*

Within column p-values: ‡ Unconc. and conc. Ultra vs. TB-LAM.

Within row p-values: *HIV-negative vs. HIV-positive.

Abbreviations: Conc., concentrated; CI, confidence interval; NPV, negative predictive value; PF, p fluid; PPV, positive predictive value; TB-LAM, Determine TB-LAM, Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Chapter 5

Xpert MTB/RIF Ultra accurately diagnoses pulmonary tuberculosis using bronchial fluid in an HIV-endemic setting

Supplementary material is attached as Appendix IV

Abstract

Background: Some patients are unable to produce sputum for the tuberculosis (TB) diagnosis, thereby requiring a bronchoscopy to obtain bronchoalveolar lavage fluid (BALF) or bronchial wash fluid (BWF). Xpert MTB/RIF Ultra (Ultra) and unstimulated interferon gamma (uIFN-γ) diagnostic performance data on BALF and BWF are limited. Moreover, the diagnostic utility of Urine-Ultra in such patients should be further investigated.

Methods: BALF or BWF from patients (>18 years) being investigated for TB (n=356) underwent programmatic 1) MGIT960 culture, 2) Xpert (later Ultra), study 3) Ultra (unconcentrated and concentrated), and uIFN-γ on BALF or BWF. Paired urine underwent study Ultra testing. Primary analyses used a microbiological reference standard (MRS).

Results: In a head-to-head comparison (n=280) of Ultra vs. Xpert on BALF/BWF, concentrated Ultra had higher sensitivity [91% (95% confidence interval: 76, 98) vs. 58% (39, 75); p=0.002] and lower specificity [77% (72, 82) vs. 96% (93, 98); p<0.001]; but sample concentration increased Ultra's non-actionable rate [4% (12/335) vs. 1% (2/354) unconcentrated BALF/BWF; p=0.005]. HIV stratus was not associated with Ultra diagnostic accuracy. In patients that had both programmatic and study concentrated Ultras done, the latter detected more TB cases [+12% (6, 17); p<0.001], indicative of programmatic laboratory room-for-improvement. uIFN-γ had poor sensitivity in BALF/BW. Urine-Ultra had a 2% (4/185) diagnostic yield.

Conclusions: Concentrated Ultra on BALF or BWF is highly sensitive and detects more TB than Xpert. uIFN-γ on BALF or BWF and urine-Ultra cannot be recommended for the diagnosis of TB in patients who are unable to produce sputum.

(249/250)

Background

The development of rapid molecular diagnostic tests has revolutionized the tuberculosis (TB) care cascade in many regions of the world[90]. The first rapid molecular test to be approved by the World Health Organization (WHO) was the Xpert MTB/RIF (Xpert; Cepheid, Sunnyvale, CA, USA), which has subsequently achieved significant market penetration[50]. Xpert MTB/RIF Ultra (Ultra) has a higher sensitivity than Xpert on sputum, though this generally comes at the cost of a lower specificity and positive predictive value[53].

The high sensitivity of Ultra makes it particularly useful for population groups who are commonly sputum test-negative, such as people living with HIV (PLHIV) in whom the sensitivity of sputum Ultra is 81% (59, 95), compared to a sensitivity of 67% (44, 86) for Xpert. In the clinical pathway for the diagnosis of pulmonary TB, patients with a high suspicion of active TB disease who are smear-negative, overall sputum test-negative, or who fall into the 'sputum scarce' category, are often referred for flexible bronchoscopy for the collection of bronchoalveolar lavage fluid (BALF) or bronchial wash fluid (BWF). To date, one study has compared the diagnostic accuracy of Xpert to Ultra on BALF in PLHIV which showed that Ultra had higher sensitivity and similar specificity[43]. Data thus remains limited.

Furthermore, several studies have shown that interferon gamma (IFN-γ), a cytokine produced by Th1 cells in response to a *Mycobacterium tuberculosis* (*M.tb*) infection, is a useful biomarker for TB diagnosis and is highly sensitive in non-sputum specimens such as pericardial and pleural fluid[69,91]. Unstimulated IFN-γ (uIFN-γ) levels in BALF and BWF for TB diagnosis has not been previously investigated and how uIFN-γ compares to Ultra for TB diagnosis remains unevaluated.

Lastly, diagnosing TB with an easily accessible fluid such as urine could prevent the need for a bronchoscopy, which is invasive and requires clinical expertise. The utility of Ultra on urine

compared to TB-LAM was shown in PLHIV with pulmonary TB when sputum was unavailable[34]. How Urine-Ultra compares to Ultra on BALF and BWF for TB diagnosis remains unknown.

Here we describe the results of a prospective diagnostic accuracy study of Ultra on BALF and BWF, performed in a high TB burden country. We compared Ultra to Xpert and uIFN-γ on BALF and BWF and evaluated Ultra on urine.

Methods and materials

Ethics statement

The study was approved by the Stellenbosch University Human Research Ethics Committee (N16/04/050) and Tygerberg Hospital (TGH; WC_2016RP15_762).

Study design

This was a prospective observational diagnostic accuracy study which compared the performance of Ultra to Xpert on BALF and BWF collected from people suspected of having TB using a microbiological reference standard (MRS). Xpert and Ultra were done while blinded to reference standard results.

Patient recruitment

356 in-patients (>18 years) with suspected pulmonary TB (based on respiratory symptoms and chest radiography) undergoing programmatic referral and investigation and requiring a bronchoscopy (due to a negative sputum test) at TGH in Cape Town, South Africa, were consecutively recruited between January 2017 and July 2022. Patients who were on TB treatment in the previous 60 days were excluded from the study.

Bronchoalveolar lavage fluid and bronchial wash collection and programmatic testing algorithm

Flexible bronchoscopy was performed under conscious sedation by a respiratory specialist. To obtain BALF, the bronchoscope was wedged in the segment of interest, sequential aliquots of 60 ml warmed sterile saline instilled and gently manually aspirated, to a maximum of 240 ml. To obtain BWF, the bronchoscope was positioned in the segment of interest without wedge, sequential aliquots of 20ml saline instilled then aspirated into a specimen trap using the wall suction apparatus, to a total volume between 60 ml and 120 ml. Where the respiratory specialist and treating clinician decided it was clinically indicated, a fluoroscopy-guided transbronchial

lung biopsy was also obtained in the same procedure. Fluid and lung biopsies were collected and sent for programmatic testing and excess fluid was obtained for study testing (**Figure 1**). Current and previous BALF, BWF or biopsies are further described in the supplement.

Definitions

Patient groups: Patients were designated definite-TB or non-TB using different reference standards. The microbiological reference standard (MRS) was used in the for primary diagnostic accuracy analyses as it's the most specific. Per the MRS, definite TBs were culture-positive on BALF/BWF or a transbronchial biopsy, and non-TBs culture-negative on BALF/BWF or transbronchial biopsies. Unclassifiable patients had no actionable MRS test (were culture contaminated or not done). Extended microbiological standard (eMRS) and composite reference standard (CRS) definitions can be found in **Supplementary Table 1.**

Other definitions: Xpert or Ultra actionable results for TB were MTB-detected and rifampicin-susceptible, rifampicin-resistant or rifampicin-indeterminate, or MTB not detected. Trace Ultra semi-quantitation results were classified as MTB detected irrespective of previous TB status. Indeterminate results were excluded from diagnostic accuracy analyses.

Programmatic testing

Bronchoalveolar lavage fluid, bronchial wash fluid and biopsy testing

Xpert, Ultra, and culture

Xpert (version 1; Cepheid, USA) was done programmatically from 25 January 2017 – 9 April 2018 by the National Health Laboratory Service (NHLS) who did Ultra (version 1) thereafter. Sample reagent (SR) buffer (1.5 ml) was added to 500 µl (3:1 ratio) of BALF/BWF or solid lung biopsy and the mixture (2 ml) used for Xpert or Ultra[57,58]. Per the algorithm (**Figure 1**; supplementary methods), BALF/BWF (5-7.5 ml) was decontaminated (with NALC-NaOH),

neutralised, centrifuged and resuspended in phosphate buffer, and 500 µl each was used for MGIT960 liquid culture and Xpert or Ultra (additional information in supplementary method).

Mycobacterium complex (MTBC) typing and drug susceptibility testing

Genotype MTBDR*plus* (Hain Lifesciences, Germany) was done on culture-positive isolates for speciation and drug susceptibility testing[76].

Study testing

Xpert and Ultra

700 µl current BALF/BWF was tested fresh with Xpert and Ultra (version 2 and later 3) using 1.4 ml SR buffer (2:1 volume ratio) as per manufacturer's instructions[57,58] for an unconcentrated test. For a concentrated Ultra, 20 ml fresh BALF/BWF was centrifuged (1711 × *g*, 10 min, room temperature), the supernatant removed until 700 µl remained, the entire resuspended pellet volume was treated with 1.4 ml SR buffer. An unconcentrated study Xpert was done if no programmatic Xpert testing was done, and sufficient volume BALF/BWF was received. An unconcentrated Ultra was always done and volume permitting, concentrated Ultra was done.

Unstimulated interferon gamma

IFN-γ levels without antigen stimulation (uIFN-γ) were measured in duplicate using filter-sterilised (0.45 and then 0.22 µm; Thermo Scientific) supernatant attained from 1.4 ml of centrifuged (20124 x *g*, 15 minutes, room temperature) BALF/BWF using the human IFN-gamma ELISAPRO kit (Mabtech, Stockholm, Sweden), following the manufacturer's instructions[77].

Urine

5-20 ml urine stored at -80 ºC [median days stored: (4, IQR: 1-7)] or recently collected (4 ºC) were centrifuged (2862 *g*, 15 min, room temperature) and processed as described above for the

BALF/BWF[58]. An unconcentrated Ultra (700 µl urine) was done when a concentrated ultra was positive or a non-actionable was recorded (**Figure 1**).

Patient treatment and follow-up

Treatment decisions were programmatic without study involvement (no study results reported for patient management unless requested by managing clinician). Attempts were made to telephonically follow-up patients at least 12 weeks after recruitment at which point general health and TB treatment initiation statuses were recorded and, if treatment started, treatment response was queried. Patients were classified as loss-to-follow-up if at least two calls were unsuccessful, and messages were unreturned for each attempt.

Statistical analysis

Differences in diagnostic accuracy metrics were calculated using proportion tests or McNemar's test as appropriate using Stata version 16.0 (StataCorp, College Station Texas, USA) and GraphPad Prism version 8.0.1 (GraphPad Software, San Diego, USA). Euler diagrams were made with InteractiVenn[60] and receiver operator curves were generated using GraphPad Prism version 8.0.1. Statistical tests also included Mann-Whitney and Spearman's coefficient using Stata version 16.0 and GraphPad Prism version 8.0.1. STARD guidelines were followed[61]. Diagnostic yield was calculated as, of people with a positive test result [study Ultra on BF (unconcentrated or concentrated), study Xpert on BF (unconcentrated or concentrated), programmatic BF culture, programmatic culture on lung biopsy, programmatic smear microscopy, culture, concentrated Ultra and concentrated Xpert-all on separate non-site-of-disease fluid, study Urine-Ultra (concentrated or unconcentrated)], the proportion positive by one specific test.

Results

Patient characteristics

From 356 patients, 8% (30/356) were unclassifiable per the MRS. Of those that could be classified, 13% (41/326) were definite TBs and 87% (285/326) were non-TBs per the MRS. Patient characteristics are described in **Table 1**.

Bronchoalveolar lavage fluid (BALF) and bronchial wash fluid (BWF)

Comparison of programmatic and study test results

7% of patients (25/356) had no culture on BALF or BWF done (**Figure 2**). Of those who had culture on BALF or BWF done, 12% (41/331) were culture-positive, of which 58% (22/38) were Xpert-positive, 76% (31/41) were study unconcentrated Ultra-positive and 84% (32/38) were study concentrated Ultra-positive (no biopsy culture was done). Of the 86% (285/331) BALF or BWF culture-negative results, 8 patients had a lung biopsy done of which 100% (7/7) were culture-negative. Of the BALF or BWF culture-negative results, 5% (15/282) were Xpert-positive, 29% (83/283) were study unconcentrated Ultra-positive and 22% (58/264) were study concentrated Ultra-positive. 2% (5/331) of BALF or BWF cultures were contaminated and were negative by Xpert and study unconcentrated Ultra while 60% (3/5) were study concentrated Ultra-positive.

uIFN-γ concentration cut-points

Optimal cut-points: Of the 356 patients enrolled, 8% (20/356) did not have uIFN-γ done. Optimal rule-in, rule-out and Youden's index cut-point concentrations for uIFN-γ on BALF and BWF to maximise diagnostic accuracy were determined using the ROC curves (**Figure 4)**. The area under the curve (AUC) for all patients was 0.624 (95% confidence interval: 0.520, 0.723). We prioritised specificity for diagnostic accuracy analyses (**Table 2**) by using a cut-point of ≥4.2 pg/ml, which corresponds to a sensitivity of 24% (13, 38) and specificity of 95%

(92, 97). When sensitivity was prioritised (≥0.0 pg/ml), sensitivity was 100% (92, 100) and specificity was 0% (0, 1). For Youden's index (≥0.1 pg/ml), sensitivity was 47% (33, 62) and specificity was 64% (55, 73). The AUC and cut-points for PLHIV and people without HIV (**Figure 4B** and **4C)** can be seen in supplementary results and **Supplementary Table 2**.

Diagnostic accuracy results

When BALF and BWF were stratified and compared, no significant differences were seen in head-to-head and non-head-to-head analyses (**Supplementary Table 3 and 4**). BALF and BWF were therefore treated as one fluid type for subsequent analyses.

Overall (head-to-head): When study unconcentrated Ultra was compared to Xpert using the MRS (n=276) (**Table 2**), study unconcentrated Ultra had similar sensitivity to Xpert [79% (95% confidence interval: 62, 91) vs. 59% (41, 75); p=0.066] but lower specificity [72% (66, 77) vs. 96% (93, 98); p<0.001]. When study concentrated Ultra was compared to study unconcentrated Ultra, sensitivity and specificity was similar [91% (76, 98) vs. 79% (62, 91); p=0.171] and [77% (71, 82) vs. 72% (66, 77); p=0.175] respectively. Using a cut-off of 4.1 pg/ml, uIFN-γ had lower sensitivity [18% (7, 35) vs. 79% (62, 91); p<0.001] and higher specificity [96% (93, 98) vs. 72% (66, 77); p<0.001] compared to study unconcentrated Ultra. Study concentrated Ultra had higher sensitivity [91% (76, 98) vs. 59% (41, 75); p=0.002] and lower specificity [77% (71, 82) vs. 96% (93, 98); p<0.001] compared to Xpert.

Conclusions were unchanged for non-head-to-head comparisons (**Supplementary Table 5).** There were no significant differences in diagnostic accuracy of tests between MRS, eMRS and CRS in head-to-head and non-head-to-head analyses (**Supplementary Table 8 and 9**).

In people without HIV: Study unconcentrated Ultra had similar sensitivity [87% (69, 96) vs. 67% (47, 83); p=0.067] and decreased specificity [73% (66, 79) vs. 96% (92, 98); p<0.001] compared to Xpert (**Table 2**). When study concentrated Ultra was compared to study unconcentrated Ultra, sensitivity was similar [93% (78, 99) vs. 87% (69, 96); p=0.389] and

specificity was similar [79% (72, 84) vs. 73% (66, 79); p=0.168]. Using a cut-off of 4.1 pg/ml, uIFN-γ had lower sensitivity [20% (8, 39) vs. 87% (69, 96); p<0.001] and higher specificity [97% (93, 99) vs. 73% (66, 79); p<0.001] compared to study unconcentrated Ultra. Study concentrated Ultra had higher sensitivity [93% (78, 99) vs. 67% (47, 83); p=0.010] and lower specificity [79% (72, 84) vs. 96% (92, 98); p<0.001] compared to Xpert. Conclusions were similar for non-head-to-head comparisons (**Supplementary Table 5)**.

In PLHIV: When study unconcentrated Ultra was compared to Xpert, study unconcentrated Ultra had similar sensitivity [25% (1, 81) vs. 0% (0, 4); p=0.285] and decreased specificity [68% (50, 82) vs. 97% (86, 100); p=0.001] (**Table 2**). When study concentrated Ultra was compared to study unconcentrated Ultra, sensitivity [75% (19, 99) vs. 25% (1, 81); p=0.157] and specificity [70% (53, 84) vs. 68% (50, 82); p=0.802] were similar. Using a cut-off of 4.1 pg/ml, uIFN-γ had similar sensitivity [0% (0, 60) vs. 25% (1, 81); p=0.285] and increased specificity [92% (78, 98) vs. 68% (50, 82); p=0.009] compared to study unconcentrated Ultra. Study concentrated Ultra had increased sensitivity [75% (19, 99) vs. 0% (0, 60); p=0.028] and decreased specificity [70% (53, 84) vs. 97% (86, 100); p=0.002] compared to Xpert. Conclusions were unchanged for non-head-to-head comparisons (**Supplementary Table 5)**.

PLHIV compared to people without HIV: For Xpert, people without HIV had increased sensitivity compared to PLHIV [67% (47, 83) vs. 0% (0, 60); p=0.011] (**Table 2**). For study unconcentrated Ultra, people without HIV had increased sensitivity compared to PLHIV [87% (69, 96) vs. 25% (1, 81); p=0.004]. The sensitivity and specificity were similar in study concentrated Ultra and uIFN-γ across HIV statuses.

Trace recategorization strategies: When unconcentrated Ultra traces were excluded, sensitivity was unchanged [-3% (-22, 17); p=0.787] and specificity increased [+19% (12, 26); p<0.001] (**Supplementary Table 6)**. When these results were rather reclassified as negative,

sensitivity was similar [-2% (-17, 13), p=0.802] and specificity increased [+12% (4, 19), p=0.002]. Similarly, when concentrated Ultra traces were excluded or reclassified as negative, sensitivity remained unchanged [-2% (-17, 13); p=0.802] and [-15% (-32, 2); p=0.100] and specificity increased [+12% (4, 19); p=0.002] and [+12% (5, 19); p<0.001] respectively (**Supplementary Table 6**).

Within patients without previous TB: When study unconcentrated Ultra was compared to Xpert, study unconcentrated Ultra had increased sensitivity [86% (67, 96) vs. 61% (41, 78); p=0.035] and decreased specificity [70% (62, 77) vs. 95% (91, 98); p<0.001]. When study concentrated Ultra was compared to study unconcentrated Ultra, sensitivity was similar [96% (82, 100) vs. 86% (67, 96); p=0.160] and specificity increased [80% (72, 86) vs. 70% (62, 77); p=0.048]. Using a cut-off of 4.1 pg/ml, uIFN-γ had lower sensitivity [18% (6, 37) vs. 86% (67, 96); p<0.001] and increased specificity [97% (92, 99) vs. 70% (62, 77); p<0.001] compared to study unconcentrated Ultra. Study concentrated Ultra had increased sensitivity [96% (82, 100) vs. 61% (41, 78); p=0.001] and decreased specificity [80% (72, 86) vs. 95% (91, 98); p<0.001] compared to Xpert. Conclusions were unchanged for non-head-to-head comparisons (**Supplementary Table 7**).

Within patients with previous TB: When study unconcentrated Ultra was compared to Xpert, study unconcentrated Ultra had similar sensitivity [50% (3, 6) vs. 50% (3, 6); p>0.999] and decreased specificity [76% (65, 84) vs. 97% (91, 99); p<0.001] (**Table 3**). When study concentrated Ultra was compared to study unconcentrated Ultra, sensitivity was similar [67% (22, 96) vs. 50% (12, 88); p=0.558] and specificity was similar [73% (63, 82) vs. 76% (65, 84); p=0.733]. Using a cut-off of 4.1 pg/ml, uIFN-γ had decreased sensitivity [17% (0, 64) vs. 50% (12, 88); p=0.221] and increased specificity [94% (88, 98) vs. 76% (65, 84); p<0.001] compared to study unconcentrated Ultra. Study concentrated Ultra had similar sensitivity [67% (22, 96) vs. 50% (12, 88); p=0.558] and decreased specificity [73% (63, 82) vs. 97% (91, 99);

p<0.001] compared to Xpert. Conclusions were similar for non-head-to-head comparisons (**Supplementary Table 7**).

Previous TB compared to non-previous TB: Study unconcentrated Ultra and study concentrated Ultra sensitivity increases in patients without previous TB [86% (67, 96) vs. 50% (12, 88); p=0.050] and [96% (82, 100) vs. 67% (22, 96); p=0.020] and specificity was similar. Sensitivity and specificity for Xpert and uIFN-γ were similar when compared between patients who had TB previously and those who did not. Conclusions were unchanged for non-head-to-head comparisons (**Supplementary Table 7)**.

Actionable Xpert and Ultra tests: When actionable results were compared for: 1) MRS, 2) programmatic concentrated and study concentrated Xpert, 3) study unconcentrated Ultra, 4) study concentrated Ultra and 5) uIFN-γ **(**rule-in cut-off 4.1 pg/ml**)**, 12% (34/276), 11% (30/276), 34% (95/276), 31% (86/276), 6% (16/276) were positive by each test (**Figure 3A**; study unconcentrated Ultra had the highest yield followed by study concentrated Ultra).

Non-actionable Xpert and Ultra tests: Unconcentrated Xpert had an initial non-actionable rate of 3% (10/299) [only includes study Xpert results; programmatic Xpert non-actionable rates unavailable], study unconcentrated Ultra had a non-actionable rate of 1% (2/354) and for study concentrated Ultra it was 4% (12/335) (**Supplementary Table 10**). Study unconcentrated Xpert had a higher non-actionable rate compared to study unconcentrated Ultra [3% (10/299) vs. 1% (2/354); p=0.008]. Study concentrated Ultra had a higher non-actionable rate compared to study unconcentrated Ultra [4% (12/335) vs. 1% (2/354); p=0.005]. Study unconcentrated Xpert and study concentrated Ultra had similar non-actionable rates [3% (10/299) vs. 4% (12/335); p=0.870]. The most common non-actionable error codes for study unconcentrated Xpert and concentrated Ultra were 5011 (Signal loss detected in the amplification curve for analyte SPC) and 2008 (Syringe pressure reading the protocol limit) respectively. After re-

testing, all Xpert and Ultra results resolved as either positive or negative (**Supplementary Table 10**).

Rifampicin (RIF) results: Both Xpert and unconcentrated Ultra correctly identified RIF-resistance in patients who had programmatic drug susceptibility testing (MTBDR*plus*) done (**Supplementary Table 11**). Concentrated Ultra incorrectly identified one TB case as RIF-resistant.

Ultra inhibition: (SPC) C_T values showed more inhibition in study concentrated Ultras than study unconcentrated Ultras [25.90 (IQR: 24.90-26.90) vs. 25.10 (24.40-26.10); p<0.001] (**Supplementary Figure 1A**).

Ultra PCR rpoB C_{Tmin} and IS6110/IS1081 C_T: An analysis of *rpo*B C_{Tmin} and IS*6110*/IS*1081* C_T values showed that study concentrated Ultra had lower *rpo*B C_{Tmin} compared to unconcentrated Ultra [23.10 (IQR: 19.30-27.60) vs. 25.10 (20.70-30.00); p=0.001] and lower IS*6110*/IS*1081* C_T [20.00 (16.50-23.65) vs. 22.15 (17.50-26.00); p<0.001] values (**Supplementary Figure 1B and 1C).**

Relationship with bacterial load: Both Xpert and study unconcentrated Ultra *rpo*B C_{Tmin} and Ultra IS*6110*/IS*1081* C_T correlated with bacillary load measured using culture time-to-positivity (TTP). After concentration, both study concentrated Ultra *rpo*B C_{Tmin} and IS*6110*/IS*1081* C_T correlated with TTP (**Supplementary Figure 2A-E**).

<u>Comparison of study unconcentrated Ultra true-positives and false-positives per the MRS</u>

More false-positives had previous TB compared to true-positives [37% (31/83) vs. 13% (4/31), p=0.012] (**Table 4**). More true-positives started TB treatment [79% (22/28) vs. 24% (18/76), p<0.001] and if a chest X-ray was done, more true-positives showed a miliary pattern on their X-ray compared to false-positives [12% (3/26) vs. 0% (0/78), p=0.002]. False-positives had higher *rpoB* C_{Tmin} and IS*6110/*IS*1081* C_T than true-positives [*rpoB* C_{Tmin} 29.80 (27.15-31.60) vs. 22.90 (19.70-28.70); p<0.001] and [IS*6110/*IS*1081* C_T 26.30 (23.60-27.80) vs. 18.40

(16.40-23.00); p<0.001]. A greater proportion of false-positives were trace-positive than true-positives [69% (59/85) vs. 13% (4/31); p<0.001]. More study unconcentrated Ultra true-positives were also Xpert-positive than false-positives [70% (21/30) vs. 14% (12/83); p<0.001] and more study unconcentrated Ultra true-positives were also programmatic concentrated Ultra-positive than false-positives [100% (27/27) vs. 27 (18/67); p<0.001]. The characteristics of false-positives per patient information is in **Supplementary Table 14**.

Study vs. programmatic Ultra results

Concordance: In patients who received both study concentrated and programmatic concentrated Ultras on BALF or BWF, 31% (75/240) were study concentrated Ultra-positive and 19% (46/240) programmatic concentrated Ultra-positive. The former detected +12% (95% confidence interval: 7, 18; p<0.001) more TB cases (**Supplementary Table 12**). In patients who received both study unconcentrated and programmatic concentrated Ultras, study unconcentrated Ultra detected [+14% (8, 20) p<0.001] more TB cases. Positivity rates in study unconcentrated and study concentrated Ultras were similar.

Urine

Diagnostic accuracy

Of concentrated urine-Ultras tested (n=56), 4% (2/56) were non-actionable and 100% (2/2) of these resolved to actionable when unconcentrated urine was tested (50% (1/2) unconcentrated urines became Ultra-positive). When urine-Ultra was compared to study unconcentrated Ultra on BALF/BWF using the MRS (*Overall, head-to-head*; n=52) (**Supplementary Table 13**), urine-Ultra had decreased sensitivity [0% (0, 46) vs. 67% (22, 96); p=0.014] and increased specificity [93% (82, 99) vs. 72% (57, 84); p=0.006]. Trends were unchanged after stratification by HIV status and in non-head-to-head comparisons.

When actionable results were compared for: 1) MRS on BALF/BWF, 2) programmatic concentrated and study concentrated Xpert on BALF/BWF, 3) study unconcentrated Ultra on BALF/BWF, 4) study concentrated Ultra on BALF/BWF and 5) concentrated and unconcentrated Ultra on urine, 7% (3/42), 7% (3/42), 33% (14/42), 38% (16/42), 7% (3/42) were positive by each test (**Figure 3B**; study unconcentrated BALF/BWF Ultra had the highest yield; Ultra on urine detected one TB case was missed by tests on BALF/BWF).

Diagnostic Yield

Overall (patients who did or did not have the test attempted)

Study Ultra (unconc. and conc.) on BF had the highest yield 89% (165/185) followed by Xpert on BF (unconc. and conc.) and culture on BF 22% (41/185) (**Supplementary Table 15**). Urine-Ultra had a low diagnostic yield of 2% (4/185). Culture on lung biopsy and tests on non-site-of-disease fluid had low yields respectively. The diagnostic yield of tests were similar when compared by HIV status.

Overall (in patients who had the test attempted)

Smear microscopy on non-site-of-disease fluid had the highest yield 100% (4/4), followed by study Ultra (unconc. and conc.) on BF 47% (165/354), Xpert (unconc. and conc.) on BF 14% (48/350) and culture on BF 13% (41/325) (**Supplementary Table 15**).

Patient treatment status at follow-up

89% (320/359) of patients were followed-up [median (IQR: 29 weeks (15-61) since recruitment] and 21% (66/320) of those had initiated TB treatment. Of these, 18% (12/66) were unclassifiable per the MRS, 43% (28/66) had been classified as definite TB and 39% (26/66) as non-TB per the MRS. Of the definite TBs on treatment, 79% (22/28) were study unconcentrated Ultra-positive, 92% (23/25) were study concentrated Ultra-positive (two were not done and one was non-actionable), 65% (17/26) were Xpert-positive (two were not done)

and 22% (6/27) were uIFN-γ positive (one was not done). For non-TBs on treatment, 65% (17/26) were study unconcentrated Ultra-positive, 45% (9/20) were study concentrated Ultra-positive (three were not done and three were non-actionable), 16% (4/25) were Xpert-positive (one was not done) and 4% (1/26) were uIFN-γ positive. Regarding the clinical status in patients who started treatment, 94% (62/66) reported treatment completion and, of these, 86% (51/59) reported feeling clinically well (3 were not reported) and 11% (7/66) died. Overall, 11% (34/318) patients were documented to have died. 9% (3/34) were exclusively study unconcentrated Ultra-positive and 15% (5/34) were exclusively study concentrated Ultra-positive; none of these patients were placed on treatment).

Discussion

Our key findings are: 1) Ultra had, compared to Xpert, higher sensitivity and lower specificity on BALF/BWF (unchanged by HIV, alternative reference standards and the use of BALF or BWF); 2) 4 in 5 Ultra "false-positives" (MRS-negative) had improved outcomes after empirical TB treatment, indicating that a positive Ultra on BALF/BWF can be used as confirmation to start TB treatment; 3) Study Ultras detected more cases than programmatic Ultras on BALF/BWF; 4) uIFN-γ on BALF/BWF is not sensitive enough to detect many TB cases, and 5) Urine-Ultra had low sensitivity but could reduce the proportion (approximately 2%) of patients with presumptive TB who require a bronchoscopy in our setting. These data show high sensitivity of Ultra on BALF/BWF for TB diagnosis.

Concentrated BF on Ultra had improved sensitivity compared to Xpert, including when Ultra trace-positives were excluded or reclassified as negative. This was similar to a study in China which showed that Ultra on BALF had a higher sensitivity compared to Xpert in people living with HIV when a reference of culture was used[43]. A systematic review comparing Ultra and Xpert on sputum also showed Ultra's superior sensitivity[53]. Notably, the sensitivity of Ultra was further improved in patients who did not have TB previously, which could be due to the absence of remnant DNA from previous TB episodes acting as inhibitors. Moreover, when unconcentrated Ultra was compared to Xpert, improved sensitivity was not observed which may be due to the paucibacillary nature of BALF/BWF or due to the low number of culture-confirmed TB cases in our study. We therefore recommend an Ultra on concentrated BF if possible (even with an increased non-actionable rate).

Both unconcentrated and concentrated Ultra had lower specificity compared to Xpert and resulted in approximately 3 in 10 MRS-negative, Ultra-positive cases (this was unchanged with alternative reference standards but improved with trace reclassification or exclusion). Previous studies of Ultra on BALF in adults[43] and children[92] (both in China) did not observe this

decreased specificity in Ultra. It is however consistent with previous diagnostic evaluations of Ultra compared to Xpert which showed that on non-sputum specimens[40,80] and sputum[81], Ultra had decreased specificity. More Ultra false-positives (MRS-negative) had previous TB than true-positives, this however was not significantly associated with specificity as seen in sputum[53]. Most Ultra 'false-positives' (most of whom were Ultra trace-positive) that started and completed TB treatment, had improved health however. Moreover, approximately 7 in 10 non-TBs were put on empirical TB treatment, which strongly suggests that these patients were true-positives but had microbiologically undetectable TB disease. Therefore, low bacterial load in this sputum-scarce population and the imperfect nature of reference standards on EPTB specimens[24,49] might be contributing to underestimation of Ultra specificity.

When programmatic and study concentrated Ultras were compared, study Ultras had a higher TB yield. This may be due to specimen processing differences in programmatic testing which decontaminates BALF/BWF before testing unlike the study, thus extracellular DNA might be lost. Additionally, more SR buffer is used in programmatic Ultras (3:1 SR to sample ratio) compared to study Ultras (2:1), diluting the specimen further. This suggests that implementing these programmatic processing procedures would improve the diagnosis of TB on BALF/BWF.

uIFN-γ has low sensitivity and high specificity in diagnosing TB when the cut-point of 4.1 pg/ml was used but misses 8 in 10 TB cases. uIFN-γ has not been previously tested in BALF/BWF for TB diagnosis. This is dissimilar to meta-analyses on pericardial fluid[69] and pleural fluid[45] that showed both high sensitivity and specificity. The low sensitivity in addition to time and laboratory labour suggests that uIFN-γ is not useful to diagnose TB.

Ultra on urine has been shown to be useful for PTB diagnosis in patients who could produce a sputum[34]. In our study, urine detected approximately 1 in 10 TB cases which are few but suggests that universal Urine-Ultra testing could reduce the need for bronchoscopies in a subset

of patients for TB diagnosis, however, this would mean that additional laboratory testing is required.

These results have strengths and limitations. Programmatic Xpert (initially done programmatically) was not always done, therefore study Xpert (which was unconcentrated) was done, resulting in two different processing methods. Additionally, BALF/BWF is programmatically decontaminated after which fluid is further diluted for Ultra testing (or Xpert initially) by more SR buffer compared to study Xpert or Ultra (which uses the recommended 2× SR buffer amount[57,58]). Programmatic Xpert and Ultra's sensitivity might therefore be underreported.

In conclusion, in a programmatic clinical setting in patients with presumptive TB, Ultra more accurately detects TB cases than Xpert regardless of HIV status, and concentrating BF increases Ultra non-actionable results. Many Ultra "false-positives", most of whom were trace-positive, had improved health after empirical treatment, showing that Ultra might be detecting cases missed by the MRS and could close a gap in the care cascade. Programmatic Ultra testing could benefit from processing optimisation. The benefits of uIFN-γ's high specificity is offset by its poor sensitivity and labour needed, indicating that it would not be practical in BALF/BWF. Furthermore, Urine-Ultra could reduce patients who require a bronchoscopy for TB diagnosis, but there remains a need for better urine-based tests. Importantly, BALF and BWF have similar diagnostic accuracy and can be used interchangeably as BF for TB diagnosis. We therefore recommend that a positive Ultra (including trace-positive) on BALF or BWF be used for TB diagnosis.

Acknowledgements

The authors thank the study participants, National Health Laboratory Services, Tygerberg Hospital, Dawn Davis, Ruth Wilson, Nandi Niemand, Roxanne Higgit, Selisha Naidoo, Dr Megan Hendrikse and Zaida Palmer. Research reported in this publication was supported by the South African Medical Research Council. GT acknowledges funding from the EDCTP2 programme supported by the European Union (RIA2018D-2509, PreFIT;RIA2018D-2493, SeroSelectTB; RIA2020I-3305, CAGE-TB) and the National Institutes of Health (D43TW010350; U01AI152087; U54EB027049; R01AI136894). CN is supported by the Fogarty International Center of the National Institutes of Health (K43TW012302). The degree from which this study emanated was made possible through funding by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the SAMRC Internship Scholarship Programme. We acknowledge the support of the South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. The content is solely the responsibility of the authors and does not necessarily represent the official views of any funders. Cepheid donated cartridges but had no role in study design or analysis.

Competing Interests

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Cepheid donated Xpert and Ultra cartridges but did not have a role in study design or result interpretation.

Funding

GT acknowledges funding from the EDCTP2 programme supported by the European Union (grant SF1401, OPTIMAL DIAGNOSIS) and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Numbers U01AI152087 and

R01AI136894. CN is supported by the Fogarty International Center of the National Institutes of Health under Award Number K43TW012302.

Figure legends

Figure 1: Specimen collection and diagnostic procedures done in participants with suspected TB. Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; TB, tuberculosis; uIFN-γ, unstimulated interferon-γ; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Figure 2: Summary of test results done programmatically (culture, concentrated Xpert, concentrated Ultra) and by the study (unconcentrated and concentrated Ultra, unconcentrated Xpert when the programme did not do Xpert) on BALF or BWF. Unconcentrated Ultra detected TB in BALF/BWF culture-negative specimens missed by concentrated Ultra and Xpert. Sample concentration decreased the Ultra positivity rate in culture-negative specimens. Data are n/N (%). Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; RIF, rifampicin; TB, tuberculosis; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

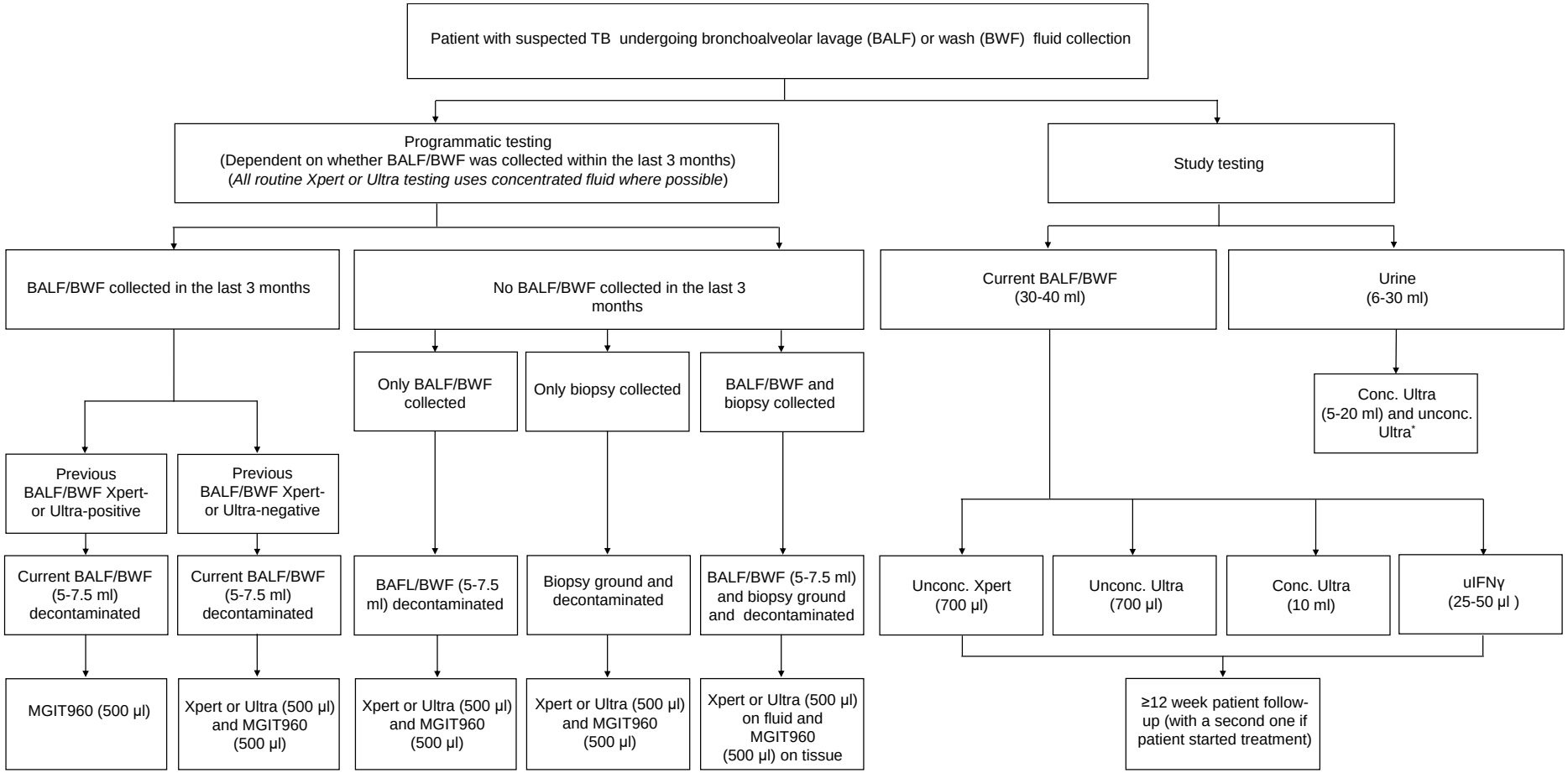
Figure 3: Euler diagrams showing positive results from different tests on BALF/BWF and/or urine versus the MRS. **(A)** MRS (programmatic culture on BALF or BWF and/or lung biopsy), Xpert (both programmatic concentrated and study unconcentrated; when programmatic testing was not done, a study unconcentrated Xpert was done), study unconcentrated Ultra, study concentrated Ultra and study uIFN-γ (cut-off 4.1 pg/ml) on BALF/BWF. Both study unconcentrated Ultra and concentrated Ultra were positive in many cases undetected by the MRS, Xpert and uIFN-γ; Xpert detected one TB case missed other tests. **(B)** MRS, Xpert (both programmatic concentrated and study unconcentrated), study unconcentrated Ultra and study concentrated Ultra on BALF/BWF and Ultra on urine (consists of both study concentrated and study unconcentrated Ultra; when a concentrated test on urine had an erroneous result, a study unconcentrated Ultra was done) in 42 patients. Study unconcentrated Ultra on BALF or BWF followed by study concentrated Ultra on BALF or BWF detected TB cases missed by other

tests; Ultra on urine detected one TB case missed by tests on BALF/BWF and MRS and Xpert on BALF/BWF detected no TB cases missed by other tests. Data are n/N (%). Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; MRS, microbiological reference standard; TB, tuberculosis; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Figure 4: Receiver operator characteristic (ROC) curve analysis of uIFN-γ on bronchoalveolar lavage fluid or bronchial wash fluid in patients with suspected TB. Values are AUCs and 95% confidence intervals. **(A)** AUC of uIFN-γ in all patients, **(B)** in people living with HIV and **(C)** people without HIV. uIFN-γ on bronchoalveolar lavage fluid or bronchial wash fluid has poor diagnostic accuracy. Abbreviations: AUC, area under the receiver operating characteristic curve; ROC, receiver operator characteristics; TB, tuberculosis; uIFN-γ, unstimulated interferon gamma.

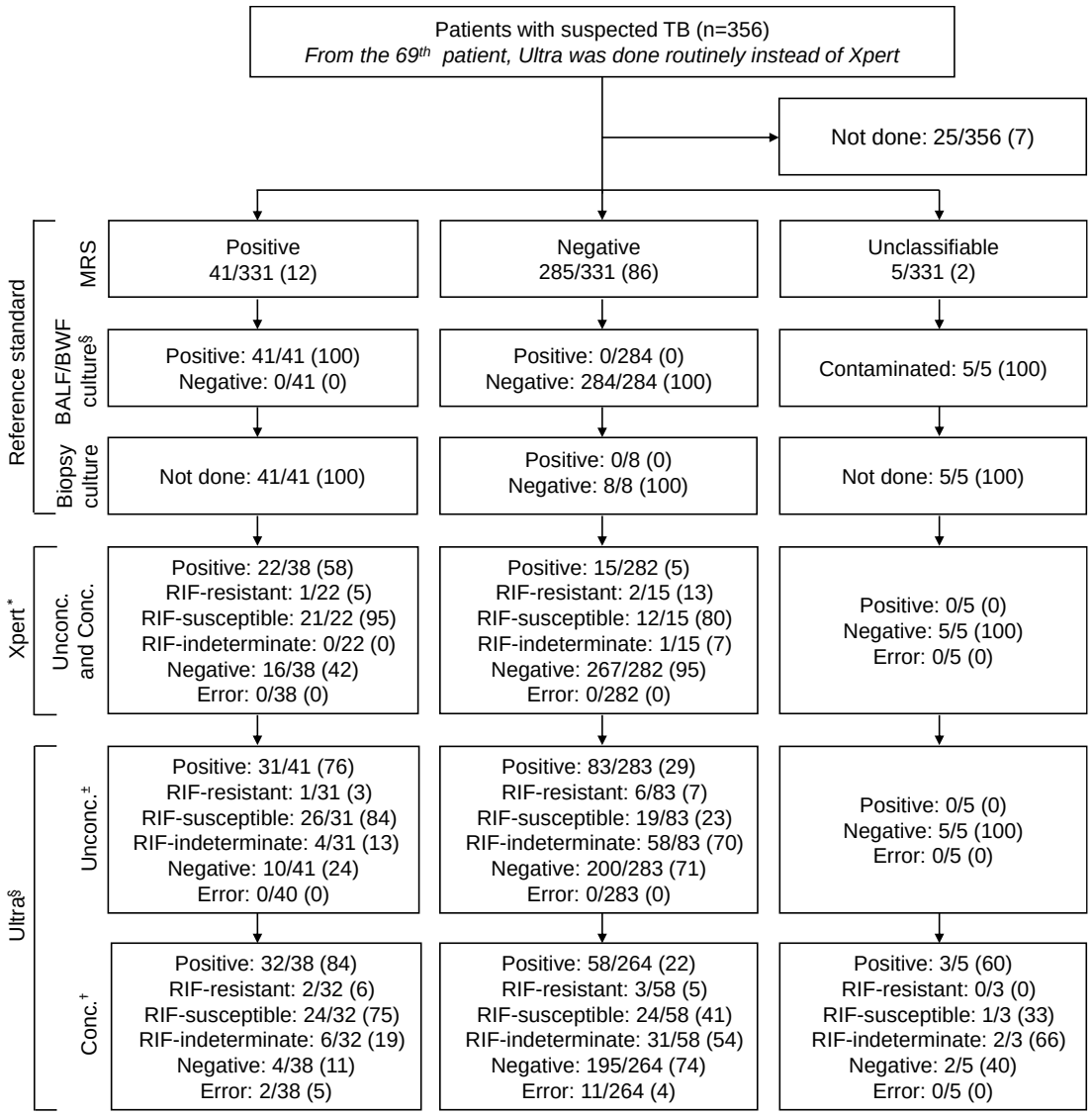
Figure 5: Sensitivity and specificity of tests on BF compared to the MRS, eMRS and CRS. uIFN-γ on BF had low sensitivity, and when compared to Xpert, Ultra had lower specificity. Abbreviations: BF, bronchial fluid; CI, confidence interval; CRS, composite reference standard; conc., concentrated; eMRS, extended microbiological standard; MRS, microbiological reference standard; uIFN-γ, unstimulated interferon-γ, Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Figure 1



*Unconc. Urine-Ultra was only done when a concentrated ultra was positive or non-actionable.

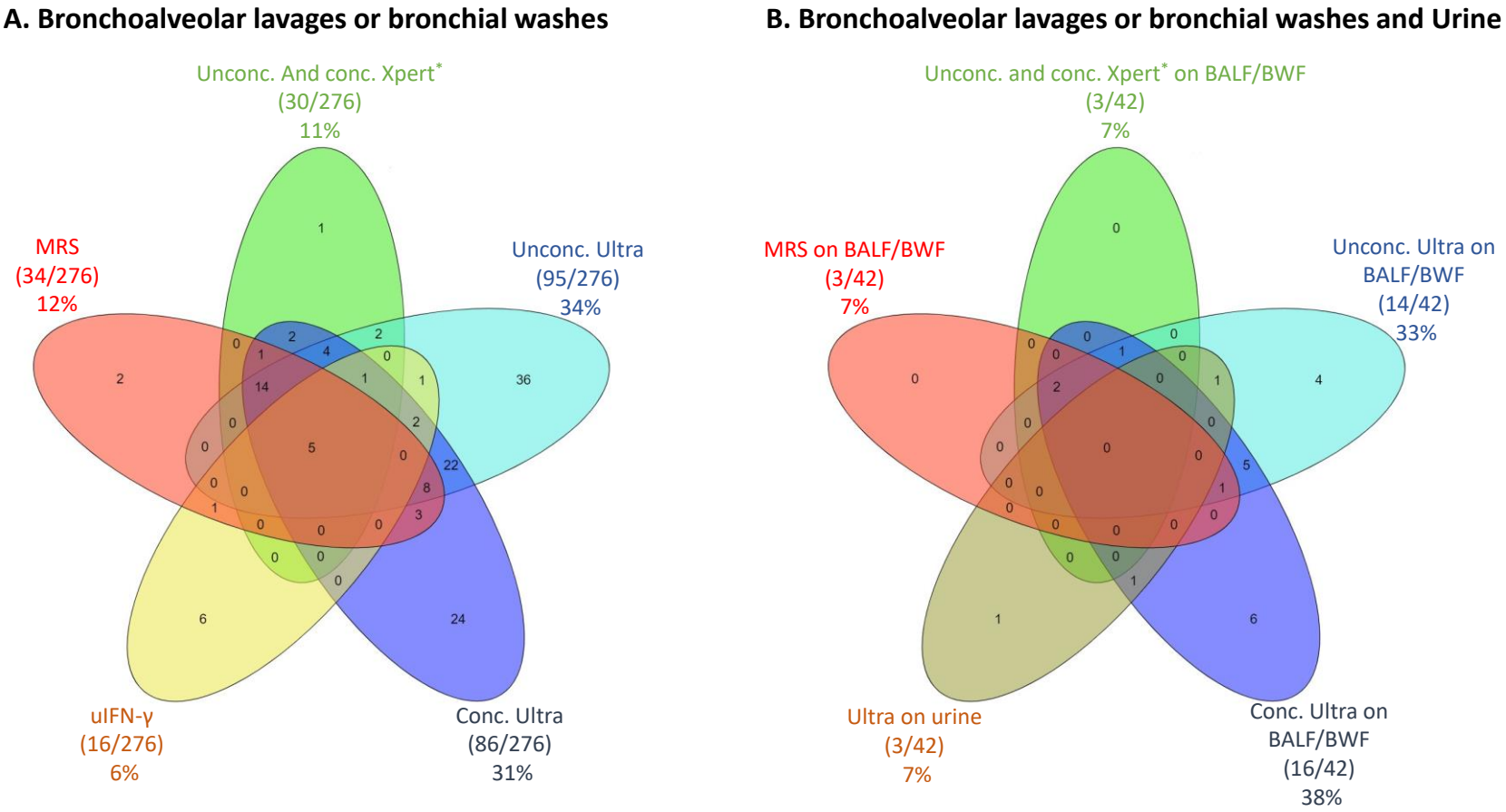
Figure 2



§27% (11/41) of BALF/BWF culture-positives had MTBDR*plus* done (one MTBDR*plus* RIF-resistant, remainder MTBDR*plus* RIF-susceptible). Both Ultra and Xpert detected the RIF-resistant culture-positive.
*Includes programmatic concentrated Xpert and study unconcentrated Xpert (latter done by the study if former not done programmatically-only BALF/BWF was used). Three MRS-positive and three MRS-negative Xperts were not done.
§Both unconcentrated and concentrated study Ultras were only done on BALF/BWF.
±Two unconcentrated Ultras were not done in patients with culture-negative BALF/BWF results.
†Study concentrated Ultras were not done in three MRS-positive patients and in 21 MRS-negatives.

.

Figure 3



*Includes programmatic concentrated Xpert and study unconcentrated Xpert (latter done by the study if former not done programmatically-only BALF/BWF was used).

Figure 4

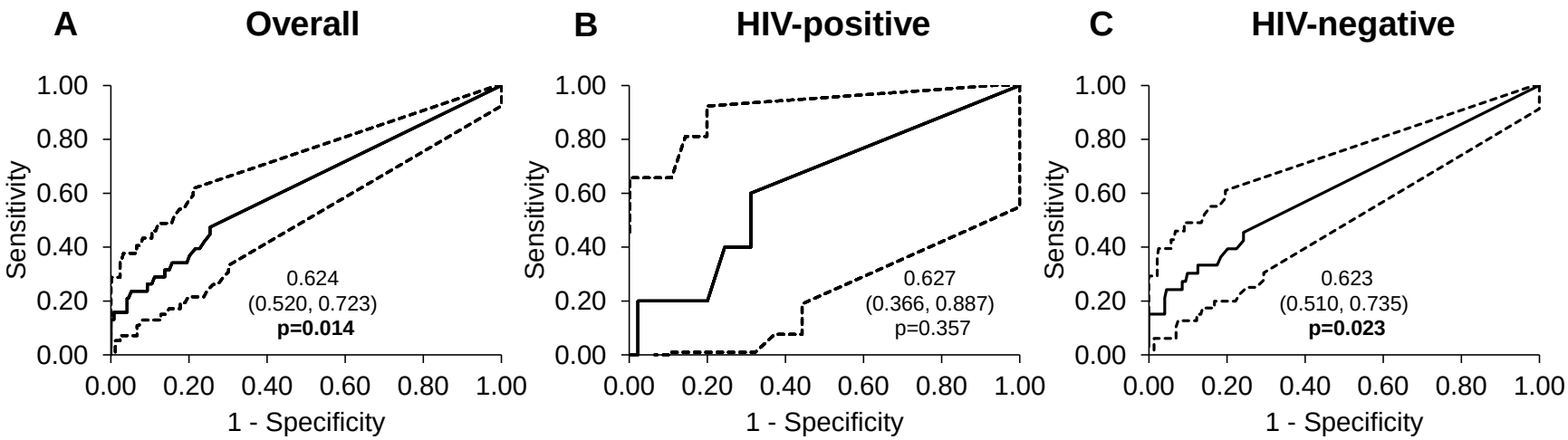


Figure 5

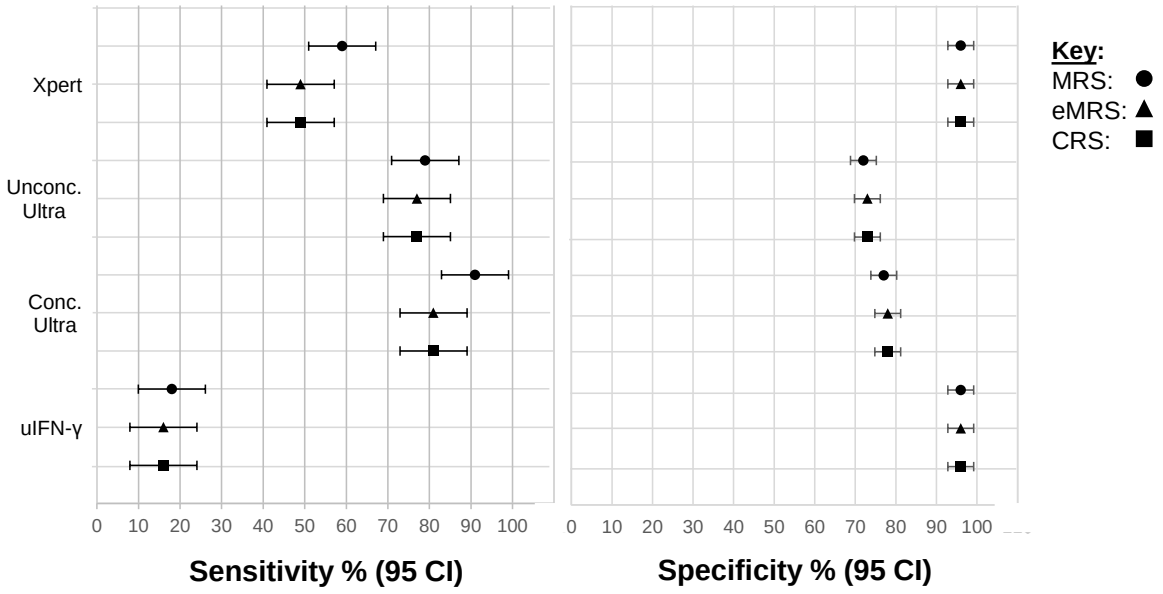


Table 1: Demographic and clinical characteristics by culture status. Non-TBs were more likely to have had previous TB and more definite-TBs experienced a cough and night sweats, showed a miliary pattern on chest X-rays, started TB treatment and had higher uIFN-γ concentrations. Data are median (IQR) or n (%).

	Overall (n=326)	**Definite-TB** (n=41)	**Non-TB** (n=285)
Demographics			
Age (years)	52 (40-62)	55 (35-64)	52 (40-62) p=0.860
Female	170/326 (52)	21/41 (51)	149/285 (52) p=0.899
Clinical characteristics			
HIV	51/326 (16)	5/41 (12)	46/285 (16) p=0.516
On ART	43/51 (84)	4/5 (80)	39/46 (85) p=0.780
CD4 (cells/µl)	337 (196-501)	309 (42-956)	355 (200-497) p=0.949
Previous TB	119/326 (37)	8/41 (20)	111/285 (39) **p=0.016**
Pulmonary TB	116/119 (97)	8/8 (100)	108/111 (97) p=0.638
Extrapulmonary TB	3/119 (3)	0/8 (0)	3/111 (3) p=0.638
Tobacco smoker	89/326 (27)	9/41 (22)	80/285 (28) p=0.459
Antibiotic use within 1 year	74/293 (25)	10/39 (26)	64/254 (25) p=0.953
Symptoms: Cough	170/325 (52)	29/40 (48)	151/285 (53) **p=0.020**
Fever	42/323 (13)	8/40 (20)	34/283 (12) p=0.160
Night sweats	70/322 (22)	14/39 (36)	56/283 (20) **p=0.022**
Weight loss	150/323 (46)	23/40 (58)	127/283 (45) p=0.134
Chest X-ray results: Normal	5/325 (2)	0/41 (0)	5/284 (2) p=0.392
Cardiomegaly	5/325 (2)	1/41 (2)	4/284 (2)

			p=0.616
Pulmonary infiltrates	214/325 (66)	27/41 (66)	187/284 (66) p=0.999
Hilar lymphadenopathy	15/325 (5)	3/41 (7)	12/284 (4) p=0.378
Miliary pattern	5/325 (2)	4/41 (10)	1/284 (0) **p<0.001**
Pleural effusion	15/325 (5)	2/41 (5)	13/284 (5) p=0.932
Unilateral cavitation	11/325 (3)	1/41 (2)	10/284 (4) p=0.720
Bilateral cavitation	4/325 (1)	1/41 (2)	3/284 (1) p=0.453
TB treatment			
TB treatment (self-reported after 12-week follow up)	54/293 (18)	28/37 (76)	26/256 (10) **p<0.001**
Study tests			
uIFN-γ	0.0 (0.0-0.70)	0.0 (0.0-3.0)	0.0 (0.0-0.5) **p=0.003**

Missing data: CD4 count, 10; antibiotic use within 1 year, 33; Cough, 1; Fever, 3; Night sweats, 4; Weight loss, 3, Chest X-ray, 1; TB treatment (unsuccessful follow-up), 33; uIFN-γ, 19.

Abbreviations: ART, antiretroviral therapy; TB, tuberculosis; uIFN-γ, unstimulated interferon gamma.

Table 2: Head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uINF-γ on BALF or BWF using the MRS stratified by HIV status. Study unconcentrated Ultra had similar sensitivity and lower specificity compared to Xpert, had similar diagnostic accuracy compared to study concentrated Ultra, and study unconcentrated had higher sensitivity and lower specificity compared to uIFN-γ. Study concentrated Ultra had higher sensitivity and lower specificity compared to Xpert. Xpert and study unconcentrated Ultra has higher sensitivity in patients that are HIV-negative. Similar patterns were observed in non-head-to-head comparisons (Supplementary Table 5). Data are %, 95% CI, and n/N.

	All patients (n=276)				**HIV-negative (n=235)**				**HIV-positive (n=41)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	59 (41, 75) 20/34	96 (93, 98) 232/242	67 (47, 83) 20/30	94 (91, 97) 232/246	67 (47, 83) 20/30	96 (92, 98) 196/205	69 (49, 85) 20/29	95 (91, 98) 196/206	0 (0, 60) 0/4 **p=0.011**	97 (86, 100) 36/37 p=0.635	0 (0, 97) 0/1 p=0.150	90 (76, 97) 36/40 p=0.199
Unconc. Ultra (study)	79 (62, 91) 27/34 p=0.066‡	72 (66, 77) 174/242 **p<0.001**‡	28 (20, 39) 27/95 **p<0.001**‡	96 (92, 98) 174/181 p=0.389‡	87 (69, 96) 26/30 p=0.067‡	73 (66, 79) 149/205 **p<0.001**‡	32 (22, 43) 26/82 **p<0.001**‡	97 (93, 99) 149/153 p=0.278‡	25 (1, 81) 1/4 p=0.285‡ **p=0.004**	68 (50, 82) 25/37 **p=0.001**‡ p=0.524	8 (0, 36) 1/13 p=0.774‡ p=0.075	89 (72, 98) 25/28 p=0.924‡ **p=0.041**
Conc. Ultra (study)	91 (76, 98) 31/34 p=0.171± **p=0.002**§	77 (71, 82) 187/242 p=0.175± **p<0.001**§	36 (26, 47) 31/86 p=0.272± **p=0.004**§	98 (95, 100) 187/190 p=0.174± **p=0.028**§	93 (78, 99) 28/30 p=0.389± **p=0.010**§	79 (72, 84) 161/205 p=0.168± **p<0.001**§	39 (28, 51) 28/72 p=0.351± **p=0.006**§	99 (96, 100) 161/163 p=0.367± p=0.051§	75 (19, 99) 3/4 p=0.157± **p=0.028**§ p=0.225	70 (53, 84) 26/37 p=0.802± **p=0.002**§ p=0.270	21 (5, 51) 3/14 p=0.315± p=0.605§ p=0.213	96 (81, 100) 26/27 p=0.317± p=0.336§ p=0.339
uIFN-γ (rule-in cut-off 4.1 pg/ml)	18 (7, 35) 6/34 **p<0.001**¥	96 (93, 98) 232/242 **p<0.001**¥	38 (15, 65) 6/16 p=0.462¥	89 (85, 93) 232/260 **p=0.008**¥	20 (8, 39) 6/30 **p<0.001**¥	97 (93, 99) 198/205 **p<0.001**¥	46 (19, 75) 6/13 p=0.306¥	89 (84, 93) 198/222 **p=0.003**¥	0 (0, 60) 0/4 p=0.285¥ p=0.324	92 (78, 98) 34/37 **p=0.009**¥ p=0.187	0 (0, 71) 0/3 p=0.620¥ p=0.137	89 (75, 97) 34/38 p=0.980¥ p=0.958

Within column p-values: ‡ Xpert vs. unconc. Ultra (study), ± Unconc. Ultra (study) vs. conc. Ultra (study), §Xpert vs. conc. Ultra (study), ¥Unconc. Ultra (study) vs. uIFN-γ in patients of the same previous TB status (overall, previous TB, no previous TB).
Within row p-values: HIV-negative vs. HIV-positive.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; CI, confidence interval; conc., concentrated; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; unconc., unconcentrated; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Table 3: Head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on BALF or BWF stratified by previous TB status using the MRS. Study unconcentrated Ultra had similar sensitivity and lower specificity compared to Xpert, had similar diagnostic accuracy compared to study concentrated Ultra, and study unconcentrated had higher sensitivity and lower specificity compared to uIFN-γ. Study concentrated Ultra had higher sensitivity and lower specificity compared to Xpert. Study unconcentrated and concentrated Ultra has higher sensitivity in patients with no previous TB. Similar patterns were observed in non-head-to-head comparisons (Supplementary Table 7). Data are % (95% CI), and n/N.

	All patients (n=276)				**No previous TB (n=180)**				**Previous TB (n=96)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	59 (41, 75) 20/34	96 (93, 98) 232/242	67 (47, 83) 20/30	94 (91, 97) 232/246	61 (41, 78) 17/28	95 (91, 98) 145/152	71 (49, 87) 17/24	93 (88, 96) 145/156	50 (12, 88) 3/6 p=0.628	97 (91, 99) 87/90 p=0.631	50 (12, 88) 3/6 p=0.333	97 (91, 99) 87/90 p=0.225
Unconc. Ultra (study)	79 (62, 91) 27/34 p=0.066‡	72 (66, 77) 174/242 **p<0.001**‡	28 (20, 39) 27/95 **p<0.001**‡	96 (92, 98) 174/181 p=0.389‡	86 (67, 96) 24/28 **p=0.035**‡	70 (62, 77) 106/152 **p<0.001**‡	34 (23, 47) 24/70 **p=0.002**‡	96 (91, 99) 106/110 p=0.234‡	50 (12, 88) 3/6 p>0.999‡ **p=0.050**	76 (65, 84) 68/90 **p<0.001**‡ p=0.330	12 (3, 31) 3/25 **p=0.034**‡ **p=0.034**	96 (88, 99) 68/71 p=0.767‡ p=0.841
Conc. Ultra (study)	91 (76, 98) 31/34 p=0.171± **p=0.002**§	77 (71, 82) 187/242 p=0.175± **p<0.001**§	36 (26, 47) 31/86 p=0.272± **p=0.004**§	98 (95, 100) 187/190 p=0.174± **p=0.028**§	96 (82, 100) 27/28 p=0.160± **p=0.001**§	80 (72, 86) 121/152 **p=0.048**± **p<0.001**§	47 (33, 60) 27/58 p=0.158± **p=0.045**§	99 (96, 100) 121/122 p=0.140± **p=0.011**§	67 (22, 96) 4/6 p=0.558± p=0.558§ **p=0.020**	73 (63, 82) 66/90 p=0.733± **p<0.001**§ p=0.261	14 (4, 33) 4/28 p=0.806± **p=0.050**§ **p=0.004**	97 (90, 100) 66/68 **p<0.001**± p=0.889§ p=0.261
uIFN-γ (rule-in cut-off 4.1 pg/ml)	18 (7, 35) 6/34 **p<0.001**¥	96 (93, 98) 232/242 **p<0.001**¥	38 (15, 65) 6/16 p=0.462¥	89 (85, 93) 232/260 **p=0.008**¥	18 (6, 37) 5/28 **p<0.001**¥	97 (92, 99) 147/152 **p<0.001**¥	50 (19, 81) 5/10 p=0.334¥	86 (80, 91) 147/170 **p=0.006**¥	17 (0, 64) 1/6 p=0.221¥ p=0.945	94 (88, 98) 85/90 **p<0.001**¥ p=0.392	17 (0, 64) 1/6 p=0.759¥ p=0.182	94 (88, 98) 85/90 p=0.700¥ **p=0.049**

Within column p-values: ‡ Xpert vs. unconc. Ultra (study), ± Unconc. Ultra (study) vs. conc. Ultra (study), §Xpert vs. conc. Ultra (study), ¥Unconc. Ultra (study) vs. uIFN-γ in patients of the same previous TB status (overall, previous TB, no previous TB).
Within row p-values: No previous TB vs. previous TB.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; CI, confidence interval; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; unconc., unconcentrated; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Table 4: Comparison of patient and microbiology characteristics study unconcentrated Ultra TP or FP status per the MRS. More FPs had TB previously, had higher *rpoB* and IS*6110*/IS*1081* C_{Ts} and were Ultra trace positive compared to TPs. More TPs showed a miliary pattern on their chest X-ray, were started on TB treatment and were Xpert- or programmatic Ultra-positive. Data are n (%) or median (IQR).

	TPs (n=31)	**FPs** (n=83)
Clinical characteristics		
HIV-positive	2/31 (6)	15/83 (18) p=0.121
CD4 count (cells/µl)	633 (309-956)	295 (111-519) p=283
ART	2/2 (100)	12/15 (80) p=0.486
Previous TB	4/31 (13)	31/83 (37) **p=0.012**
Antibiotic use within 1 year	8/30 (27)	20/76 (26) p=0.971
Current smoker	6/31 (19)	24/83 (29) p=0.302
Symptoms: Cough	13/30 (43)	49/83 (59) p=0.139
Fever	4/30 (13)	12/81 (15) p=0.844
Night sweats	11/30 (37)	19/82 (23) p=0.153
Weight loss	17/30 (57)	39/82 (48) p=0.393
Chest X-ray results: Normal	0/26 (0)	1/78 (1) p=0.562
Cardiomegaly	0/26 (0)	1/78 (1) p=0.562
Pulmonary infiltrates	22/26 (85)	56/78 (72) p=0.191
Hilar lymphadenopathy	1/26 (4)	3/78 (4) p>0.999
Miliary pattern	3/26 (12)	0/78 (0) **p=0.002**
Pleural effusion	2/26 (8)	5/78 (6)

		p=0.821
Unilateral cavitation	1/26 (4)	3/78 (4) p>0.999
Bilateral cavitation	0/26 (0)	0/78 (0) p>0.999
TB treatment		
¥Patients initiated on TB treatment after 12-week follow-up	22/28 (79)	18/76 (24) **p<0.001**
If on treatment, did the patient report improved health?	17/21 (81)	13/16 (81) p=0.982
Study unconcentrated Ultra result information		
rpoB C_{Tmin}	22.90 (19.70-28.70)	29.80 (27.15-31.60) **p<0.001**
IS*6110*/IS*1081* C_T	18.40 (16.40-23.00)	26.30 (23.60-27.80) **p<0.001**
Trace semi-quantitation category	4/31 (13)	58/85 (68) **p<0.001**
SPC C_T	25.10 (24.10-28.30)	25.30 (24.50-26.10) p=0.846
Programmatic and study Xpert and programmatic Ultra information		
Positive Xpert (unconc. and conc.)	21/30 (70)	12/83 (14) **p<0.001**
Positive programmatic conc. Ultra	27/27 (100)	18/67 (27) **p<0.001**
Study tests		
uIFN-γ	0 (0-4.20)	0 (0-1.30) p=0.109

Missing data: CD4 count, n=1; ART, n=1; antibiotic use within 1 year, n=9; cough, n=1; fever, n=3; night sweats, n=2; weight loss, n=2; Chest X-ray result, n=1; patients who were lost to follow-up, n=9; patient report of improved health, n=2 ;study unconcentrated Ultra SPC C_T excluded because it's zero, n=1; Xperts not done, n=1; programmatic concentrated Ultras not done, n=20; uIFN-γ, n=4.

Abbreviations: conc., concentrated; FP, false-positive; IS*6110*/IS*1081* C_T, cycle threshold value for the Xpert MTB/RIF Ultra IS*6110*/IS*1081* probe; *rpoB* C_{Tmin}, minimum cycle threshold value for the Xpert MTB/RIF Ultra *rpoB* probes; TP, true-positive; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

¥Study Ultra results were not reported for potential patient management.

Chapter 6

Discussion

In 2021, only 60% of those with TB were diagnosed (6.4 million of the estimated 10.6 million people)[93]. Ultra is endorsed for TB diagnosis on sputum but at the advent of the study, more data was needed regarding the usefulness Ultra on different non-sputum specimens, particularly in HIV-endemic settings. Furthermore, the impact of sample processing methods such as sample concentration of non-sputum specimens (in specimens with sufficient volume) on the diagnostic accuracy and moreover, non-actionable rate of Ultra remains unclear. Lastly, it was largely unknown how Ultra directly compares to other tests on site-of-disease fluid or non-site-of-disease fluid.

Firstly (Chapter 2), we found that in a routine clinical setting in patients with presumptive TBL and FNABs collected, Ultra detects more TBL cases than Xpert and results in more people placed on treatment. This is largely due to the inclusion of Ultra trace-positive results, but this comes with lower specificity. Ultra sensitivity and specificity on FNABs were not significantly associated with different reference standards or HIV. Study Ultra detected more TBL cases to that of programmatic Ultras when both tests were done, which indicates that programmatic Ultra testing could be optimised and thus TBL diagnosis would be improved in routine laboratories. Moreover, FNAB Ultra false-negative results are associated with PCR inhibition. Urine-Ultra had low sensitivity but detected TBL in one patient missed by tests on site-of-disease fluid, indicating that Urine-Ultra could reduce FNAB collection in a subset of patients for TBL diagnosis.

In patients with presumptive TB pericarditis (Chapter 3), we found that in those with HIV, Ultra on unconcentrated pericardial fluid had higher sensitivity and lower specificity overall compared to Xpert (alternate reference standards did not improve sensitivity). Additionally, exclusion of unconcentrated Ultra trace results improved specificity without large sensitivity decrements unlike reclassifying trace results as negative. The use of concentrated pericardial fluid on Ultra resulted in higher specificity compared to unconcentrated pericardial fluid but

this came with an increase in non-actionable results. This shows that with sufficient fluid volume and laboratory capacity, concentrated pericardial fluid Ultra should be done. The high sensitivity of uIFN-γ is offset by poor specificity, and high costs, indicating that Ultra is the better diagnostic test. The sensitivity of Urine-Ultra and TB-LAM were low but could reduce the need for pericardiocentesis for TB pericarditis diagnosis in 4% of patients, highlighting the potential of tests on urine. Our study thus demonstrates that Ultra on pericardial fluid is the superior test for TB pericarditis diagnosis.

When we evaluated tests on pleural fluid (Chapter 4), Ultra had similar sensitivity, but higher diagnostic yield compared to Xpert, and exclusion or reclassification of trace results to negative did not increase sensitivity. HIV status was not significantly associated with Ultra diagnostic accuracy, and the use of alternate reference did not significantly increase sensitivity. Ultra on unconcentrated pleural fluid has lower specificity compared to Xpert, but specificity increases when pleural fluid is concentrated by centrifugation, but this increases non-actionable results. This phenomenon was also seen in pericardial fluid and could be due to increased PCR inhibitors in these fluids preventing false-positives as they are now true-negatives, thereby increasing specificity. uIFN-γ has high sensitivity and moderate specificity on pleural fluid, suggesting that laboratories with the funds and infrastructure needed could use uIFN-γ concentration for TB pleuritis diagnosis as it detects approximately one in ten cases missed by microbiological tests on pleural fluid. TB-LAM on urine could reduce the need for a thoracentesis in 6% of patients for TB pleuritis diagnosis, particularly in people living with HIV. We therefore recommend Ultra on pleural fluid for TB pleuritis diagnosis and if laboratories have the capacity, we recommend uIFN-γ.

Lastly in bronchial fluid that included BALF and BWF (Chapter 5), we found that the diagnostic accuracy for BALF and BWF were not significantly different when compared, and thus, the fluids can be used interchangeably for TB diagnosis. Ultra on concentrated bronchial

fluid had higher sensitivity and lower specificity when compared to Xpert (HIV and alternate reference standards did not significantly change Ultra's sensitivity and specificity). 4 in 5 patients that were MRS-negative Ultra-positive, or Ultra "false-positive" started empirical treatment, indicating that Ultra could be detecting TB missed by culture. Additionally, programmatic Ultra testing would benefit from processing optimisation as seen by study Ultra detecting more TB cases. Furthermore, uIFN-γ's high specificity in bronchial fluid is offset by its poor sensitivity, which indicates impracticality for TB diagnosis using bronchial fluid. Urine-Ultra had low sensitivity but detected TB in 2% of patients with presumptive TB. Our study therefore recommends that a positive Ultra (including trace-positive) on bronchial fluid be used for TB diagnosis.

Even though these fluids were from different cohorts of patients, we noted that when contrasting Ultra across fluids, Ultra had the highest significant sensitivity increase in FNABs followed by pericardial fluid in PLHIV when compared to Xpert. Moreover, all fluids had decreased specificity compared to Xpert, and notably after sample concentration, specificity increased in pericardial and pleural fluid, and not in bronchial fluid. uIFN-γ performed the best in pleural fluid, and urinary Ultra and TB-LAM were universally not sensitive in these cohorts.

The findings in this study highlights Ultra's high sensitivity and moderate specificity in patients with presumptive TBL, TB pericarditis, and TB pleuritis and PTB. We recommend that a positive Ultra with the inclusion of trace results be used for EPTB and PTB diagnosis.

Table 1: Summary of whether a diagnostic test is recommended or not recommended per chapter and its' pertaining condition. For each condition, we recommend that a urine-test first be done and if negative, an Ultra on site-of-disease should be done, after which treatment should commence if positive.

Condition	**Diagnostic test**					
	Unconc. Ultra on site-of-disease fluid	**Conc. Ultra on site-of-disease fluid**	**Xpert on site-of-disease fluid**	**uIFN-γ on site-of-disease fluid**	**Conc. Ultra on urine**	**TB-LAM on urine**
TBL (Chapter 2)	Recommended	N/A	Not recommended	N/A	Recommended	N/A
TB pericarditis (Chapter 3)	Recommended	Recommended	Not recommended	Not recommended	Recommended	Recommended
TB pleuritis (Chapter 4)	Recommended	Recommended	Not recommended	Recommended	Recommended	Recommended
Sputum-scarce PTB (Chapter 5)	Recommended	Recommended	Not recommended	Not recommended	Recommended	N/A

Abbreviations: Conc., concentrated; TB-LAM, Determine TB-LAM; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF; uIFN-γ, unstimulated interferon-γ.

Chapter 7

Conclusion and future work

In conclusion, the studies involved in this dissertation have provided much needed diagnostic accuracy data for Ultra on non-sputum specimens in a HIV-endemic setting of Cape Town, South Africa. We have shown the high sensitivity of Ultra, thereby showing that a positive Ultra result should be used for TBL, TB pericarditis, TB pleuritis and PTB diagnosis. We have also shown that Ultra detects more TB than its' predecessor, Xpert on FNABs, pericardial fluid, pleural fluid, and bronchial fluid. Moreover, we showed that when pericardial and pleural are concentrated by centrifugation, specificity of Ultra increases. Furthermore, our study showed that uIFN-γ has high sensitivity in pleural fluid for TB pleuritis diagnosis, but does not provide significant benefit over Ultra. Finally, we showed that urine tests could potentially reduce invasive sampling and associated delays in a subset of patients. Moreover, concentrated urine Ultra testing will detect more TB than unconcentrated urine, but this comes with increased non-actionable results. Thus, there remains a need for more sensitive tests. These novel findings have the potential to reduce the diagnostic gaps in patients with EPTB or sputum scarce PTB and could inform local and international diagnostic policy guidelines.

For future work, more studies should investigate Ultra 'false-positive' results on site-of-disease fluid, which would provide more insight as to Ultra's lower specificity compared to Xpert. The reason for Ultra's increased specificity when certain fluids are concentrated by centrifugation remains unknown and should thus be further investigated. Additionally, more sensitive tests should be evaluated using non-site-disease fluid such as urine or blood (C-reactive protein and RNA biosignatures) in these non-sputum cohorts. This could provide valuable data on why tests on site-of-disease fluid including culture and Ultra might not be detecting TB at the site-of-disease.

Chapter 8

Additional academic outputs (in order of date published)

Table 2: Additional academic outputs, not directly counting towards the PhD threshold but which the candidate had intellectual input into during their PhD (appendix number), with publication year, journal (impact factor), and take-home messages.

Manuscript title (Appendix #)	Publication year	Authorship	Journal (impact factor)	Candidate's role	Take home messages
1. Extract from used Xpert MTB/RIF Ultra cartridges is useful for accurate second-line drug-resistant tuberculosis diagnosis with minimal rpoB-amplicon cross-contamination risk 10.1038/s41598-020-59164-3 (V)	2020	Joint 1st	*Scientific Reports* (4.99)	Processing of Ultra cartridges, extracting of Ultra cartridges and editing of manuscript.	• MTBDR*sl* on Ultra diamond cartridge extract (CE) is concordant with sputum drug-susceptibility testing (DST) results when a threshold of C_{Tmin} <25 is applied and risk of *rpoB* cross-contamination is unlikely. • 16S rRNA qPCR, MTBDR*plus*, MTBDR*sl* were feasible on other cartridge chambers. • MTBDR*plus* and FluoroType was not feasible on Ultra and Xpert diamond cartridge extracts.
2. Frequent suboptimal thermocycler ramp rate usage negatively impacts GenoType MTBDRsl VER 2.0 performance for second-line drug resistant tuberculosis diagnosis https://doi.org/10.1016/j.jmoldx.2022.01.003 (VI)	2022	7th	*Journal of Molecular Diagnostics* (5.34)	Assisted with the blind readings of MTBDR*sl* strips.	• In sputa, valid results improved by 21% when using the optimal ramp rate on MTBDR*sl.* • MTBDR*sl* banding call and drug susceptibility call reader disagreement worsened at the suboptimal ramp rate. • Laboratory respondents that corrected their ramp rate reported fewer non-valid results on smear-negative specimens.
3. More than Mycobacterium tuberculosis: site-of-disease microbial communities, and their functional and clinical profiles in tuberculous lymphadenitis http://dx.doi.org/10.1136/thorax-2022-219103 (VII)	2022	7th	*Thorax* (9.20)	Assisted with study co-ordination and management and reviewed the manuscript.	• TBL at the site-of-disease is not microbially homogeneous. • Distinct microbial community clusters exist that, in our setting, are associated with different clinical characteristics, and immunomodulatory potentials. • Non-Mycobacterium-dominated dTBL lymphotypes, which contain taxa potentially targeted by TB treatment, were associated with milder, potentially earlier stage disease.

References

1. Gupta RK, Lawn SD, Bekker L-G, Caldwell J, Kaplan R, Wood R. Impact of human immunodeficiency virus and CD4 count on tuberculosis diagnosis: analysis of city-wide data from Cape Town, South Africa. *The international journal of tuberculosis and lung disease* 2013; **17**(8): 1014-22.
2. Chakravorty S, Simmons AM, Rowneki M, et al. The new Xpert MTB/RIF Ultra: improving detection of Mycobacterium tuberculosis and resistance to rifampin in an assay suitable for point-of-care testing. *MBio* 2017; **8**(4): e00812-17.
3. Denkinger CM, Schumacher SG, Boehme CC, Dendukuri N, Pai M, Steingart KR. Xpert MTB/RIF assay for the diagnosis of extrapulmonary tuberculosis: a systematic review and meta-analysis. *European Respiratory Journal* 2014; **44**(2): 435-46.
4. World Health Organization. Global tuberculosis report, 2022.
5. World Health Organization. Global tuberculosis report, 2020.
6. Sharma S, Mohan A, Kadhiravan T. HIV-TB co-infection: epidemiology, diagnosis & management. *Indian Journal of Medical Research* 2005; **121**(4): 550-67.
7. Leeds IL, Magee MJ, Kurbatova EV, et al. Site of extrapulmonary tuberculosis is associated with HIV infection. *Clinical Infectious Diseases* 2012; **55**(1): 75-81.
8. Sharma S, Mohan A. Extrapulmonary tuberculosis. *Indian Journal of Medical Research* 2004; **120**: 316-53.
9. Tortoli E, Russo C, Piersimoni C, et al. Clinical validation of Xpert MTB/RIF for the diagnosis of extrapulmonary tuberculosis. *European Respiratory Journal* 2012; **40**(2): 442-7.
10. Ligthelm LJ, Nicol MP, Hoek KG, et al. Xpert MTB/RIF for rapid diagnosis of tuberculous lymphadenitis from fine-needle-aspiration biopsy specimens. *Journal of clinical microbiology* 2011; **49**(11): 3967-70.
11. Balaji J, Sundaram SS, Rathinam SN, Rajeswari PA, Kumari MV. Fine needle aspiration cytology in childhood TB lymphadenitis. *The Indian Journal of Pediatrics* 2009; **76**(12): 1241-6.
12. Pandie S, Peter JG, Kerbelker ZS, et al. Diagnostic accuracy of quantitative PCR (Xpert MTB/RIF) for tuberculous pericarditis compared to adenosine deaminase and unstimulated interferon-γ in a high burden setting: a prospective study. *BMC medicine* 2014; **12**(1): 101.
13. Mayosi BM, Wiysonge CS, Ntsekhe M, et al. Mortality in patients treated for tuberculous pericarditis in sub-Saharan Africa. *South African Medical Journal* 2008; **98**(1): 36-40.
14. Krishna R, Antoine MH, Rudrappa M. Pleural Outflow.
15. Friedrich SO, von Groote-Bidlingmaier F, Diacon AH. Xpert MTB/RIF assay for diagnosis of pleural tuberculosis. *Journal of clinical microbiology* 2011; **49**(12): 4341-2.
16. Shaw JA, Diacon AH, Koegelenberg CF. Tuberculous pleural effusion. *Respirology* 2019; **24**(10): 962-71.
17. Khalil KF, Butt T. Diagnostic yield of bronchoalveolar lavage gene Xpert in smear-negative and sputum-scarce pulmonary tuberculosis. *J Coll Physicians Surg Pak* 2015; **25**(2): 115-8.
18. Mayosi BM, Burgess LJ, Doubell AF. Tuberculous pericarditis. *Circulation* 2005; **112**(23): 3608-16.
19. Bisognin F, Lombardi G, Lombardo D, Re MC, Dal Monte P. Improvement of Mycobacterium tuberculosis detection by Xpert MTB/RIF Ultra: A head-to-head comparison on Xpert-negative samples. *PloS one* 2018; **13**(8): e0201934.

20. Bahr NC, Nuwagira E, Evans EE, et al. Diagnostic accuracy of Xpert MTB/RIF Ultra for tuberculous meningitis in HIV-infected adults: a prospective cohort study. *The Lancet infectious diseases* 2018; **18**(1): 68-75.
21. Zeka AN, Tasbakan S, Cavusoglu C. Evaluation of the GeneXpert MTB/RIF assay for rapid diagnosis of tuberculosis and detection of rifampin resistance in pulmonary and extrapulmonary specimens. *Journal of clinical microbiology* 2011; **49**(12): 4138-41.
22. Ngabonziza JCS, Ssengooba W, Mutua F, et al. Diagnostic performance of smear microscopy and incremental yield of Xpert in detection of pulmonary tuberculosis in Rwanda. *BMC infectious diseases* 2016; **16**(1): 660.
23. Chakravorty S, Sen MK, Tyagi JS. Diagnosis of extrapulmonary tuberculosis by smear, culture, and PCR using universal sample processing technology. *Journal of clinical microbiology* 2005; **43**(9): 4357-62.
24. Kohli M, Schiller I, Dendukuri N, et al. Xpert MTB/RIF Ultra and Xpert MTB/RIF assays for extrapulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews* 2021; (1).
25. Liao C-H, Chou C-H, Lai C-C, et al. Diagnostic performance of an enzyme-linked immunospot assay for interferon-γ in extrapulmonary tuberculosis varies between different sites of disease. *Journal of Infection* 2009; **59**(6): 402-8.
26. Pai M, Denkinger CM, Kik SV, et al. Gamma interferon release assays for detection of Mycobacterium tuberculosis infection. 2014; **27**(1): 3-20.
27. World Health Organization. Use of tuberculosis interferon-gamma release assays (IGRAs) in low- and middle income countries. Policy statement. 2011.
28. Jiang J, Shi H-Z, Liang Q-L, Qin S-M, Qin X-J. Diagnostic value of interferon-γ in tuberculous pleurisy: a metaanalysis. *Chest* 2007; **131**(4): 1133-41.
29. Dheda K, van Zyl-Smit R, Sechi LA, et al. Utility of quantitative T-cell responses versus unstimulated interferon-γ for the diagnosis of pleural tuberculosis. *European Respiratory Journal* 2009; **34**(5): 1118-26.
30. Lawn SD. Point-of-care detection of lipoarabinomannan (LAM) in urine for diagnosis of HIV-associated tuberculosis: a state of the art review. *BMC infectious diseases* 2012; **12**(1): 1-12.
31. Broger T, Koeppel L, Huerga H, et al. Diagnostic yield of urine lipoarabinomannan and sputum tuberculosis tests in people living with HIV: a systematic review and meta-analysis of individual participant data. *The Lancet Global Health* 2023; **11**(6): e903-e16.
32. Atherton RR, Cresswell FV, Ellis J, et al. Detection of Mycobacterium tuberculosis in urine by Xpert MTB/RIF Ultra: a useful adjunctive diagnostic tool in HIV-associated tuberculosis. *International Journal of Infectious Diseases* 2018; **75**: 92-4.
33. Cresswell FV, Ellis J, Kagimu E, et al. Standardized urine-based tuberculosis (TB) screening with TB-lipoarabinomannan and Xpert MTB/RIF Ultra in Ugandan adults with advanced human immunodeficiency virus disease and suspected meningitis. Open forum infectious diseases; 2020: Oxford University Press US; 2020. p. ofaa100.
34. Andama A, Jaganath D, Crowder R, et al. Accuracy and incremental yield of urine Xpert MTB/RIF Ultra versus Determine TB-LAM for diagnosis of pulmonary tuberculosis. *Diagnostic microbiology and infectious disease* 2020; **96**(1): 114892.
35. Peter JG, Theron G, Muchinga TE, Govender U, Dheda K. The diagnostic accuracy of urine-based Xpert MTB/RIF in HIV-infected hospitalized patients who are smear-negative or sputum scarce. *PloS one* 2012; **7**(7): e39966.
36. Antel K, Oosthuizen J, Malherbe F, et al. Diagnostic accuracy of the Xpert MTB/Rif Ultra for tuberculosis adenitis. *BMC infectious diseases* 2020; **20**: 1-8.

37. Aggarwal AN, Agarwal R, Dhooria S, Prasad KT, Sehgal IS, Muthu V. Xpert MTB/RIF Ultra versus Xpert MTB/RIF for diagnosis of tuberculous pleural effusion: A systematic review and comparative meta-analysis. *Plos one* 2022; **17**(7): e0268483.
38. Meldau R, Randall P, Pooran A, et al. Same-day tools, including Xpert Ultra and IRISA-TB, for rapid diagnosis of pleural tuberculosis: a prospective observational study. *Journal of clinical microbiology* 2019; **57**(9): e00614-19.
39. Perez-Risco D, Rodriguez-Temporal D, Valledor-Sanchez I, Alcaide F. Evaluation of the Xpert MTB/RIF ultra assay for direct detection of Mycobacterium tuberculosis complex in smear-negative extrapulmonary samples. *Journal of clinical microbiology* 2018; **56**(9): e00659-18.
40. Wu X, Tan G, Gao R, et al. Assessment of the Xpert MTB/RIF Ultra assay on rapid diagnosis of extrapulmonary tuberculosis. *International Journal of Infectious Diseases* 2019; **81**: 91-6.
41. Banada PP, Koshy R, Alland D. Detection of Mycobacterium tuberculosis in blood by use of the Xpert MTB/RIF assay. *Journal of clinical microbiology* 2013; **51**(7): 2317-22.
42. Patel VB, Theron G, Lenders L, et al. Diagnostic accuracy of quantitative PCR (Xpert MTB/RIF) for tuberculous meningitis in a high burden setting: a prospective study. *PLoS medicine* 2013; **10**(10): e1001536.
43. Zhang P, Liu H, Wang H, et al. Performance of Xpert MTB/RIF Ultra for the Diagnosis of Pulmonary Tuberculosis Using Bronchoalveolar Lavage Samples in People Living with HIV/AIDS (PLWHA) in China: A Prospective Study. *HIV/AIDS (Auckland, NZ)* 2021; **13**: 905.
44. Pandie S, Peter JG, Kerbelker ZS, et al. Diagnostic accuracy of quantitative PCR (Xpert MTB/RIF) for tuberculous pericarditis compared to adenosine deaminase and unstimulated interferon-γ in a high burden setting: a prospective study. *BMC medicine* 2014; **12**(1): 1-11.
45. Aggarwal AN, Agarwal R, Dhooria S, Prasad KT, Sehgal IS, Muthu V. Unstimulated pleural fluid interferon gamma for diagnosis of tuberculous pleural effusion: a systematic review and meta-analysis. *Journal of clinical microbiology* 2021; **59**(5): e02112-20.
46. Chien J-Y, Lin C-K, Yu C-J, Hsueh P-R. Usefulness of Xpert MTB/RIF Ultra to rapidly diagnose sputum smear-negative pulmonary tuberculosis using bronchial washing fluid. *Frontiers in Microbiology* 2020; **11**: 588963.
47. Derese Y, Hailu E, Assefa T, et al. Comparison of PCR with standard culture of fine needle aspiration samples in the diagnosis of tuberculosis lymphadenitis. *The Journal of Infection in Developing Countries* 2012; **6**(01): 53-7.
48. Gautam H, Agrawal SK, Verma SK, Singh UB. Cervical tuberculous lymphadenitis: clinical profile and diagnostic modalities. *International journal of mycobacteriology* 2018; **7**(3): 212.
49. Dheda K, Ruhwald M, Theron G, Peter J, Yam WC. Point-of-care diagnosis of tuberculosis: Past, present and future. *Respirology* 2013; **18**(2): 217-32.
50. Albert H, Nathavitharana RR, Isaacs C, Pai M, Denkinger CM, Boehme CC. Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: what lessons have we learnt and how can we do better? *Eur Respir J* 2016; **48**(2): 516-25.
51. Schnippel K, Meyer-Rath G, Long L, Stevens W, Sanne I, Rosen SJSAMJ. Diagnosing Xpert MTB/RIF-negative TB: Impact and cost of alternative algorithms for South Africa. 2013; **103**(2): 101-6.
52. García-Basteiro AL, DiNardo A, Saavedra B, et al. Point of care diagnostics for tuberculosis. *Pulmonology* 2018; **24**(2): 73-85.
53. Mishra H, Reeve BW, Palmer Z, et al. Xpert MTB/RIF Ultra and Xpert MTB/RIF for diagnosis of tuberculosis in an HIV-endemic setting with a high burden of previous

tuberculosis: a two-cohort diagnostic accuracy study. *The Lancet Respiratory Medicine* 2020; **8**(4): 368-82.
54. Naidoo P, Theron G, Rangaka MX, et al. The South African tuberculosis care cascade: estimated losses and methodological challenges. *The Journal of infectious diseases* 2017; **216**(suppl_7): S702-S13.
55. Wright CA, Bamford C, Prince Y, et al. Mycobacterial transport medium for routine culture of fine needle aspiration biopsies. *Archives of disease in childhood* 2010; **95**(1): 48-50.
56. World Health Organization. Global tuberculosis report. . 2019.
57. World Health Organization. Xpert MTB/RIF implementation manual: Technical and operational 'how-to': practical considerations. 2014.
58. Cepheid. Xpert MTB/RIF Ultra Product Brochure. 2017.
59. Girdler-Brown B, Dzikiti LN. Hypothesis tests for the difference between two population proportions using Stata. *Southern African Journal of Public Health* 2018; **2**(3): 63-8.
60. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim RJBb. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. 2015; **16**(1): 169.
61. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ open* 2016; **6**(11): e012799.
62. Theron G, Peter J, Calligaro G, et al. Determinants of PCR performance (Xpert MTB/RIF), including bacterial load and inhibition, for TB diagnosis using specimens from different body compartments. *Scientific reports* 2014; **4**(1): 1-10.
63. Spener-Gomes R, Costa AG, de Melo HF, et al. Examination of respiratory specimens improves microbiological diagnosis of patients with presumptive extrapulmonary tuberculosis. *International Journal of Infectious Diseases* 2021.
64. Theron G, Venter R, Smith L, et al. False-positive Xpert MTB/RIF results in retested patients with previous tuberculosis: frequency, profile, and prospective clinical outcomes. *Journal of clinical microbiology* 2018; **56**(3).
65. Theron G, Venter R, Calligaro G, et al. Xpert MTB/RIF results in patients with previous tuberculosis: can we distinguish true from false positive results? *Clinical Infectious Diseases* 2016; **62**(8): 995-1001.
66. Ganchua SKC, White AG, Klein EC, Flynn JLJPp. Lymph nodes—The neglected battlefield in tuberculosis. 2020; **16**(8): e1008632.
67. Bulterys MA, Wagner B, Redard-Jacot M, et al. Point-of-care urine LAM tests for tuberculosis diagnosis: a status update. *Journal of clinical medicine* 2020; **9**(1): 111.
68. Reuter H, Burgess L, Doubell A. Epidemiology of pericardial effusions at a large academic hospital in South Africa. *Epidemiology & Infection* 2005; **133**(3): 393-9.
69. Liu C, Cui Y-L, Ding C-M, et al. Diagnostic accuracy of interferon-gamma in pericardial effusions for tuberculous pericarditis: a meta-analysis. *Journal of thoracic disease* 2018; **10**(2): 854.
70. Bastian A, Meissner A, Lins M, Siegel E, Möller F, Simon R. Pericardiocentesis: differential aspects of a common procedure. *Intensive care medicine* 2000; **26**(5): 572-6.
71. Abdulgader SM, Okunola AO, Ndlangalavu G, et al. Diagnosing Tuberculosis: What Do New Technologies Allow Us to (Not) Do? *Respiration* 2022; **101**(9): 797-813.
72. Nathavitharana RR, Garcia-Basteiro AL, Ruhwald M, Cobelens F, Theron G. Reimagining the status quo: How close are we to rapid sputum-free tuberculosis diagnostics for all? *EBioMedicine* 2022: 103939.
73. Andrianto A, Mertaniasih NM, Gandi P, et al. Diagnostic test accuracy of Xpert MTB/RIF for tuberculous pericarditis: a systematic review and meta-analysis. *F1000Research* 2020; **9**.

74. Denkinger CM, Kalantri Y, Schumacher SG, et al. Challenges in the development of an immunochromatographic interferon-gamma test for diagnosis of pleural tuberculosis. *PLoS One* 2013; **8**(12): e85447.
75. Theron G, Peter J, Dowdy D, Langley I, Squire SB, Dheda K. Do high rates of empirical treatment undermine the potential effect of new diagnostic tests for tuberculosis in high-burden settings? *The Lancet infectious diseases* 2014; **14**(6): 527-32.
76. Hain Lifescience. GenoType MTBDRplus VER 2.0. Instructions for use. 2015.
77. Mabtech. Human IFN-gamma ELISAPRO kit Datasheet and Protocol. 2016.
78. Abbott. Determine TB LAM Ag package insert. 2020.
79. Cummings P, McKnight B. Analysis of matched cohort data. *The Stata Journal* 2004; **4**(3): 274-81.
80. Minnies S, Reeve BW, Rockman L, et al. Xpert MTB/RIF Ultra Is Highly Sensitive for the Diagnosis of Tuberculosis Lymphadenitis in a High-HIV Setting. *Journal of clinical microbiology* 2021; **59**(12): e01316-21.
81. Osei Sekyere J, Maphalala N, Malinga LA, Mbelle NM, Maningi NE. A comparative evaluation of the new genexpert MTB/RIF ultra and other rapid diagnostic assays for detecting tuberculosis in pulmonary and extra pulmonary specimens. *Scientific Reports* 2019; **9**(1): 1-9.
82. Pandie S, Peter JG, Kerbelker ZS, et al. The diagnostic accuracy of pericardial and urinary lipoarabinomannan (LAM) assays in patients with suspected tuberculous pericarditis. *Scientific reports* 2016; **6**(1): 1-9.
83. Székely R, Sossen B, Mukoka M, et al. Multicentre accuracy trial of FUJIFILM SILVAMP TB LAM test in people with HIV reveals lot variability. *medRxiv* 2022.
84. Shaw JA, Irusen EM, Diacon AH, Koegelenberg CF. Pleural tuberculosis: a concise clinical review. *The clinical respiratory journal* 2018; **12**(5): 1779-86.
85. Vorster MJ, Allwood BW, Diacon AH, Koegelenberg CF. Tuberculous pleural effusions: advances and controversies. *Journal of thoracic disease* 2015; **7**(6): 981.
86. Cantey EP, Walter JM, Corbridge T, Barsuk JH. Complications of thoracentesis: incidence, risk factors, and strategies for prevention. *Current opinion in pulmonary medicine* 2016; **22**(4): 378.
87. Sehgal IS, Dhooria S, Aggarwal AN, Behera D, Agarwal R. Diagnostic performance of Xpert MTB/RIF in tuberculous pleural effusion: systematic review and meta-analysis. *Journal of clinical microbiology* 2016; **54**(4): 1133-6.
88. Meldau R, Peter J, Theron G, et al. Comparison of same day diagnostic tools including Gene Xpert and unstimulated IFN-γ for the evaluation of pleural tuberculosis: a prospective cohort study. *BMC pulmonary medicine* 2014; **14**(1): 1-10.
89. Yu W, Shen Y, Zhu P, Chen D. Head-to-head comparison of the efficacy of Xpert MTB/RIF Ultra and Xpert MTB/RIF for the diagnosis of tuberculous pleurisy: A systematic review and meta-analysis. *Medicine* 2022; **101**(20).
90. Naidoo P, Theron G, Rangaka MX, et al. The South African Tuberculosis Care Cascade: Estimated Losses and Methodological Challenges. *J Infect Dis* 2017; **216**(suppl_7): S702-S13.
91. Shaw JA, Koegelenberg CF. Pleural tuberculosis. *Clinics in Chest Medicine* 2021; **42**(4): 649-66.
92. Sun L, Zhu Y, Fang M, et al. Evaluation of Xpert MTB/RIF ultra assay for diagnosis of childhood tuberculosis: a multicenter accuracy study. *Journal of Clinical Microbiology* 2020; **58**(9): e00702-20.
93. Treatment Action Group. Tuberculosis diagnostics pipeline report, 2022.

Appendix I

Xpert MTB/RIF Ultra is highly sensitive for the diagnosis of tuberculosis lymphadenitis in a high-HIV setting

Chapter 2

(Supplementary material)

Table of Contents

Supplementary Table 1: Reference standard definitions

	MRS*	eMRS†	CRS‡
	Site of disease fluid		
MGIT960 Culture	✓	✓	✓
Cytology	✓	✓	✓
	Non-site-of disease fluid		
Smear microscopy	×	✓	✓
Xpert	×	✓	
Ultra	×	✓	✓
MGIT960 Culture	×	✓	✓
	Treatment information		
TB treatment initiated	×	×	✓
Response to treatment self-reported by patient	×	×	✓
	Case definitions		
Reference standard **positive (Definite TB cases)**	Any MRS test positive	Any eMRS test positive	Any eMRS test positive/or TB treatment was initiated and response to treatment documented
Reference standard **negative (Non-TB patients)**	No MRS test positive	No eMRS test positive	No eMRS test positive and patient not initiated on treatment
Probable TB patients	N/A	N/A	No eMRS test positive, but treatment initiated
Unclassifiable	No positive MRS test and site-of-disease fluid culture contaminated or not done	No positive eMRS test and site-of-disease fluid culture contaminated or not done	No positive eMRS test and site-of-disease fluid culture contaminated or not done, or treatment not initiated

Abbreviations: CRS, composite reference standard; eMRS, extended microbiological reference standard; MGIT960 culture, Mycobacteria Growth Indicator Tube 960; Microbiological reference standard, MRS; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Definitions

Microbiological reference standard

For the microbiological reference standard (MRS), a definite TB case was defined as a fine needle aspirate (FNAB) being culture-positive or cytology-positive and a non-TB patient was defined as being FNAB culture and cytology negative. Patients were unclassifiable if they had no positive MRS test and the site-of-disease culture was either contaminated or not done or cytology was not done.

Extended microbiological reference standard

For the extended microbiological reference standard (eMRS), a definite TB case was defined as a FNAB or any other body fluid being culture-, smear-, routine Xpert- or Ultra- positive and a non-TB case was defined as FNAB and other body fluids being culture-, smear-, Xpert-and Ultra- negative. Patients were considered unclassifiable if they had no positive eMRS test and the site-of-disease culture was either contaminated or not done.

Composite reference standard

For the composite reference standard (CRS), a definite TB case was defined as a FNAB or any other body fluid being culture-, smear-, Xpert- or Ultra- positive or TB treatment was initiated and response to treatment is documented; a probable-TB case was defined as a FNAB or any other body fluid being culture-, smear-, Xpert- or Ultra- positive or the patient being initiated on TB treatment after the 12-week follow up; and a non-TB case was defined as FNAB and other body fluids being culture-, smear-, Xpert-and Ultra- negative, and the patient was not initiated on TB treatment, and the patient was diagnosed with an alternative disease. Patients were unclassifiable if they had no positive eMRS test and the site-of-disease culture was either contaminated or not done, and treatment was not initiated.

Supplementary Results

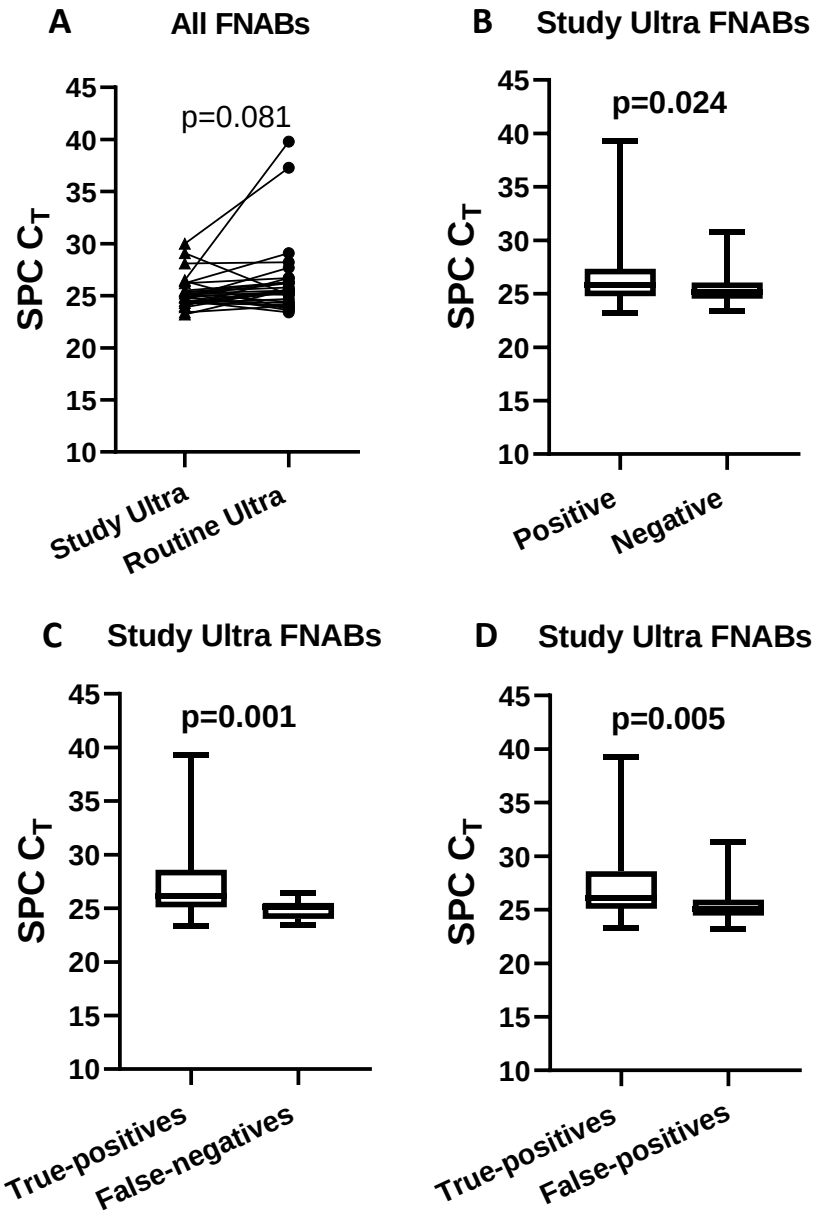
Bacterial load in study Ultra and routine Xpert

No correlations were observed between study Ultra quantitation (IS*6110*/IS*1081* C_T and *rpoB* C_{Tmin}) and culture time-to-positivity (TTP) and routine Xpert quantitation (*rpoB* C_{Tmin}) and culture TTP (**Supplementary Figure 2**).

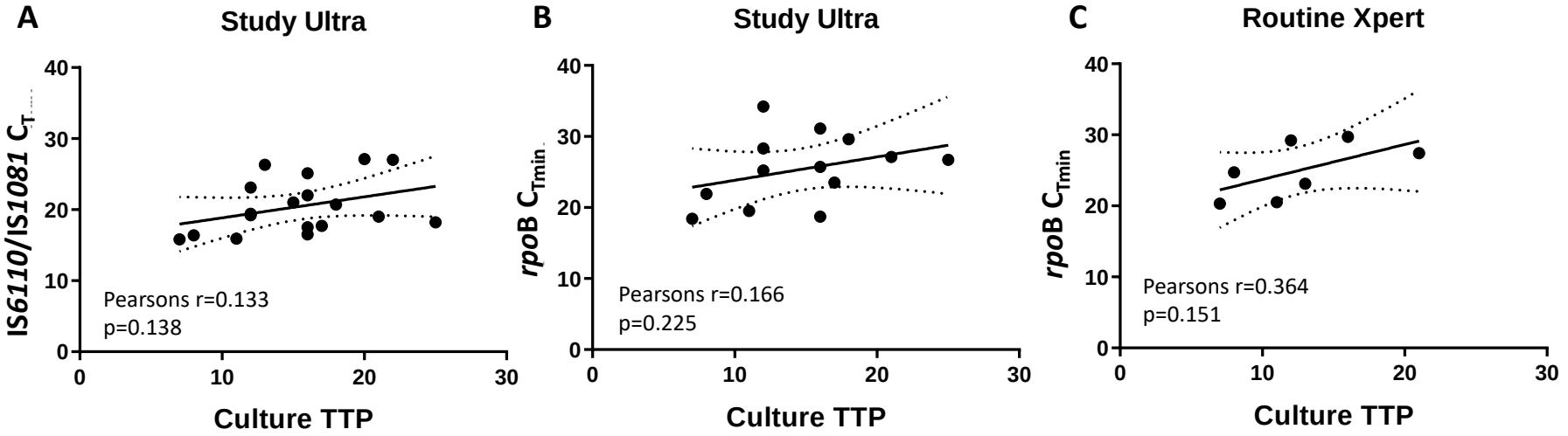
Drug susceptibility results of study Ultras on FNABs

Of 74 study Ultra-positive patients, 70% (52/74) were rifampicin-susceptible, 4% (3/74) resistant, and 26% (19/74) indeterminate (all trace). In patients who had actionable study Ultra and culture results (n=84), 20% (17/84) had MTBDR*plus* done. Of these, 18% (3/17) were MTBDR*plus* rifampicin-resistant and 82% (14/17) susceptible. 33% (1/3) of these MTBDR*plus* rifampicin-resistant patients were study Ultra rifampicin-resistant (one study Ultra trace-positive, rifampicin indeterminate and the other study Ultra negative), and 57% (8/14) of MTBDR*plus* rifampicin-susceptible patients were study Ultra rifampicin-susceptible (the remaining four were study Ultra trace-positive, rifampicin indeterminate and other remaining two were study Ultra-negative).

Supplementary Figure 1: Spaghetti and box and whiskers plots showing FNAB Ultra SPC C_T. (A) Study SPC C_T vs. routine SPC C_T. (B) SPC C_T from positive and negative study Ultras. More inhibition was observed in positives. (C) SPC C_T in true-positive vs. false-negative study Ultras, showing that greater inhibition is associated with Ultra missing TBL cases. Abbreviations: FNAB, fine needle aspirate biopsy; SPC C_T; sample processing control cycle threshold value for the Xpert MTB/RIF Ultra (Ultra) internal positive control which measures PCR inhibition; Ultra, Xpert MTB/RIF Ultra.



Supplementary Figure 2: FNAB Quantitative information of Ultra (IS*6110*/IS*1081*, *rpo*B) and Xpert (*rpo*B) compared with bacillary load (MGIT960 liquid culture TTP)**. (A)** Study Ultra IS*6110*/IS*1081* C_T vs. MGIT960 liquid culture TTP. (B) Study Ultra *rpo*B C_{Tmin} vs. MGIT960 liquid culture TTP**. (C)** Routine Xpert *rpo*B $C_T$min vs. MGIT960 liquid culture TTP. No correlations were observed. Only two culture-positive, routine Ultra-positive FNABs were present and routine Ultra results are hence not graphed. Abbreviations: FNAB, fine needle aspirate biopsy; TTP, culture time-to-positivity; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.



Supplementary Table 2: Non-head-to-head and head-to-head diagnostic accuracy analyses of Xpert and Ultra using a MRS, eMRS and CRS for the detection of Mycobacterium tuberculosis complex DNA. Conclusions were like those for the MRS **(Table 2)**. Data are %, 95% CI, and n/N

	Non-head-to-head											
	MRS				**eMRS**				**CRS**			
	n=96				**n=97**				**n=97**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert	73 (58, 85) 35/48	92 (80, 98) 44/48	90 (76, 97) 35/39	77 (64, 87) 44/57	69 (55, 81) 36/52 p=0.685*	91 (79, 98) 41/45 p=0.924*	90 (76, 97) 36/40 p=0.970*	72 (58, 83) 41/57 p=0.519*	65 (52, 77) 39/60 p=0.635± p=0.379¥	97 (86, 100) 36/37 p=0.244± p=0.274¥	98 (87, 100) 39/40 p=0.166± p=0.157¥	63 (49, 76) 36/57 p=0.317± p=0.102¥
	n=130				**n=131**				**n=131**			
Ultra	85 (73, 93) 51/60 p=0.121‡	69 (56, 79) 48/70 **p=0.003**‡	70 (58, 80) 51/73 **p=0.018**‡	84 (76, 97) 48/57 p=0.343‡	83 (71, 91) 53/64 p=0.085‡ p=0.741*	69 (56, 79) 46/67 **p=0.005**‡ p=0.991*	72 (60, 81) 53/74 **p=0.024**‡ p=0.815*	81 (68, 90) 46/57 p=0.271‡ p=0.622*	76 (65, 85) 58/76 p=0.147‡ p=0.345± p=0.207¥	71 (57, 82) 39/55 **p=0.001**‡ p=0.788± p=0.778¥	78 (67, 87) 58/74 **p=0.006**‡ p=0.343± p=0.238¥	68 (55, 80) 39/57 p=0.554‡ p=0.132± **p=0.047**¥
	Head-to-head											
	n=92				**n=92**				**n=92**			
Xpert	72 (57, 84) 33/46	93 (82, 99) 43/46	92 (78, 98) 33/36	77 (64, 87) 43/56	67 (52, 80) 33/49 p=0.642*	93 (81, 99) 40/43 p=0.068*	92 (78, 98) 33/36 p>0.999*	71 (58, 83) 40/56 p=0.518*	64 (50, 76) 35/55 p=0.691± p=0.387¥	97 (86, 100) 36/37 p=0.382± p=0.419¥	97 (85, 100) 35/36 p=0.304± p=0.304¥	64 (50, 77) 36/56 p=0.418± p=0.147¥
Ultra	91 (79, 98) 42/46 **p=0.016**‡	76 (61, 87) 35/46 **p=0.020**‡	79 (66, 89) 42/53 p=0.114‡	90 (76, 97) 35/39 p=0.105‡	88 (75, 95) 43/49 **p=0.016**‡ p=0.573*	77 (61, 88) 33/43 **p=0.035**‡ p=0.942*	81 (68, 91) 43/53 p=0.167‡ p=0.808*	85 (69, 94) 33/39 p=0.134‡ p=0.498*	84 (71, 92) 46/55 **p=0.017**‡ p=0.551± p=0.252¥	81 (65, 92) 30/37 **p=0.025**‡ p=0.636± p=0.583¥	87 (75, 95) 46/53 p=0.091‡ p=0.427± p=0.301¥	77 (61, 89) 30/39 p=0.188‡ p=0.389± p=0.129¥

Within rows: *MRS vs. eMRS, eMRS vs. CRS±, MRS vs. CRS¥; Within columns: Xpert vs. Ultra‡

Abbreviations: CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, Negative predictive value; PPV, Positive predictable value; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 3: Per patient information for study Ultra-positive patients that were MRS -negative (culture- and cytology-negative) with information on their Ultra semi-quantitation category, previous TB status, TB treatment initiation status and patient's status after at least 12-weeks of follow-up. Data are n/N (%).

Patient ID	**Previous TB**	**Study Ultra semi-quantitation category**	**Routine PCR result**	**Treatment initiated after 12-week follow up**	**Did the patient get better?** (asked telephonically if patient started treatment)
FNAB038	No	Trace	Xpert-negative	No	N/A
FNAB060	No	Very Low	Xpert-negative	Yes	Yes
FNAB072	No	Very Low	Xpert-negative	No	N/A
FNAB076	No	Medium	Xpert-positive (Low)	Yes	Yes
FNAB110	No	Trace	Xpert-negative	No	N/A
FNAB114	No	Medium	Xpert-positive (Medium)	No	N/A
FNAB128	No	Trace	Xpert-negative	No	N/A
FNAB132	No	Trace	Xpert-negative	No	N/A
FNAB172	No	Low	Xpert-negative	Yes	Yes
FNAB180	Yes	Low	Xpert-negative	No	N/A
FNAB200	No	Low	Xpert-positive (Low)	Yes	Yes
FNAB206	Yes	Trace	Not done	No	N/A
FNAB210	Yes	Trace	Ultra-negative	No	N/A
FNAB214	No	Trace	Ultra-negative	No	N/A
FNAB218	Yes	Trace	Ultra-negative	No	N/A
FNAB220	No	Trace	Ultra-positive (Very low)	Yes	Yes
FNAB230	No	Trace	Ultra-negative	No	N/A
FNAB232	Yes	Medium	Ultra-positive (Medium)	Yes	Yes
FNAB245	No	Trace	Ultra-negative	No	N/A
FNAB251	Yes	Trace	Ultra-negative	No	N/A
FNAB273	No	Trace	Ultra-negative	No	N/A
FNAB403	No	Medium	Ultra-positive (Medium)	No	N/A
Overall	6/22 (27)	Trace: 13/22 (59) Very low: 2/22 (9) Low: 3/22 (14) Medium: 4/22 (18)	Xpert: 11/21 (52) positive, 3/11 (27) negative, 8/11 (73) Ultra: 10/21 (48) positive, 3/10 (30) negative, 7/10 (70)	6/22 (27)	6/6 (100)

Missing data: Routine PCR not done, n=1.
Abbreviations: FNAB, fine needle aspirate biopsy; PCR, polymerase chain reaction; Ultra, Xpert MTB/RIF Ultra. If the patient was not initiated on TB treatment, N/A was recorded in the last column.

Supplementary Table 4: Diagnostic accuracy of Ultra on urine or FNABs measured using the MRS in a head-to-head analysis stratified by HIV status. Urine-Ultra had lower sensitivity than FNAB Ultra but increased specificity **(Table 2)**. Data are %, 95% CI, and n/N

	All patients				**HIV-negative**				**HIV-positive**			
	n=76				**n=18/75 (24)**				**n=57/75 (76)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Urine-Ultra	18 (7, 35) 6/33	98 (88, 100) 42/43	86 (42, 100) 6/7	61 (48, 72) 42/69	0 (0, 52) 0/5	100 (75, 100) 13/13	0/0	72 (47, 90) 13/18	21 (8, 41) 6/28 p=0.252*	97 (82, 100) 28/29 p=498*	86 (42, 100) 6/7	56 (41, 70) 28/50 p=0.228*
Study FNAB-Ultra	91 (76, 98) 30/33 **p<0.001‡**	60 (44, 75) 26/43 **p<0.001‡**	64 (49, 77) 30/47 p=0.252‡	90 (73, 98) 26/29 **p=0.005‡**	80 (28, 99) 4/5 **p=0.010‡**	38 (14, 68) 5/13 **p=0.001‡**	33 (10, 65) 4/12	83 (36, 100) 5/6 p=0.586‡	93 (76, 99) 26/28 **p<0.001‡** p=0.357*	69 (49, 85) 20/29 **p=0.005‡** p=0.063*	74 (57, 88) 26/35 p=0.517‡ **p=0.011***	91 (71, 99) 20/22 **p=0.004‡** p=0.595*

Missing data: Non-actionable Ultras (n=1), no HIV (n=1) in the head-to-head table.

Within column p-values: ‡ Urine-Ultra vs. FNAB-Ultra

Within row p-values: *HIV-negative vs. HIV-positive

Abbreviations: CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; Xpert, Xpert MTB/RIF; Ultra, Xpert MTB/RIF Ultra.

Supplementary Table 5: Non-head-to-head and head-to-head diagnostic accuracy analyses of Xpert and Ultra using a MRS, eMRS and CRS for the detection of *Mycobacterium tuberculosis complex* DNA with and without exclusion of Ultra trace results. Routine Xpert results were compared to study Ultra results. Ultra has similar diagnostic accuracy compared to Xpert after trace positive exclusion. Study Ultra results with trace excluded were like study Ultra results. Similar trends are seen across reference standards. Data are %, 95% CI, and n/N.

	Non-head-to-head											
	MRS				**eMRS**				**CRS**			
	n=96				**n=97**				**n=97**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert$^{\omega}$	73 (58, 85) 35/48	92 (80, 98) 44/48	90 (76, 97) 35/39	77 (64, 87) 44/57	69 (55, 81) 36/52 p=0.685*	91 (79, 98) 41/45 p=0.924*	90 (76, 97) 36/40 p=0.510*	72 (58, 83) 41/57 p=0.519*	65 (52, 77) 39/60 p=0.635$^{\pm}$ p=0.379$^{¥}$	97 (86, 100) 36/37 p=0.244$^{\pm}$ p=0.274$^{¥}$	98 (87, 100) 39/40 p=0.166$^{\pm}$ p=0.157$^{¥}$	63 (49, 76) 36/57 p=0.317$^{\pm}$ p=0.102$^{¥}$
	n=111				**n=112**				**n=112**			
Ultra excluding trace	83 (71, 92) 45/54 p=0.201‡	84 (72, 93) 48/57 p=0.248‡	83 (71, 92) 45/54 p=0.379‡	84 (72, 93) 48/57 p=0.343‡	81 (69, 90) 47/58 p=0.151‡ p=0.751*	85 (73, 93) 46/54 p=0.368‡ p=0.887*	85 (73, 94) 47/55 p=0.510‡ p=0.760*	81 (68, 90) 46/57 p=0.271‡ p=0.622*	74 (62, 84) 51/69 p=0.272‡ p=0.341$^{\pm}$ p=0.210$^{¥}$	91 (78, 97) 39/43 p=0.224‡ p=0.413$^{\pm}$ p=0.340$^{¥}$	93 (82, 98) 51/55 p=0.304‡ p=0.221$^{\pm}$ p=0.130$^{¥}$	68 (55, 80) 39/57 p=0.554‡ p=0.132$^{\pm}$ **p=0.047**$^{¥}$
Δ Trace excluded$^{\phi}$	-2 (-15, 12) p=0.808§	+15 (1, 30) **p=0.041**§	+13 (-1, 28) p=0.081§	0 (-13, 13) p>0.999§	-2 (-15, 12) p=0.799§	+16 (2, 31) **p=0.034**§	+13 (-0.03, 28) p=0.063§	0 (-14, 14) p>0.999§	-2 (-16, 12) p=0.738§	+20 (5, 35) **p=0.016**§	+15 (3, 26) **p=0.026**§	0 (-17, 17) p>0.999§
	Head-to-head											
	n=82				**n=82**				**n=82**			
Xpert$^{\omega}$	80 (64, 91) 32/40	93 (81, 99) 39/42	91 (77, 98) 32/35	83 (69,92) 39/47	74 (59, 86) 32/43 p=0.545*	92 (79, 98) 36/39 p=0.925*	91 (77, 98) 32/35 p>0.999*	77 (62, 88) 36/47 p=0.441*	69 (55, 82) 34/49 p=0.593$^{\pm}$ p=0.255$^{¥}$	97 (84, 100) 32/33 p=0.390$^{\pm}$ p=0.431$^{¥}$	97 (85, 100) 34/35 p=0.303$^{\pm}$ p=0.303$^{¥}$	68 (53, 81) 32/47 p=0.356$^{\pm}$ p=0.093$^{¥}$
Ultra excluding trace	90 (76, 97) 36/40 p=0.210‡	83 (69, 93) 35/42 p=0.178‡	84 (69, 93) 36/43 p=0.311‡	90 (76, 97) 35/39 p=0.367‡	86 (72, 95) 37/43 p=0.176‡ p=0.580*	85 (69, 94) 33/39 p=0.288‡ p=0.875*	86 (72, 95) 37/43 p=0.459‡ p=0.763*	85 (69, 94) 33/39 p=0.353‡ p=0.498*	82 (68, 91) 40/49 p=0.159‡ p=0.567$^{\pm}$ p=0.266$^{¥}$	91 (76, 98) 30/33 p=0.302‡ p=0.421$^{\pm}$ p=0.338$^{¥}$	93 (81, 99) 40/43 p=0.412‡ p=0.291 p=0.178$^{¥}$	77 (61, 89) 30/39 p=0.363‡ p=0.389$^{\pm}$ p=0.129$^{¥}$
Δ Trace excluded$^{\phi}$	-1 (-14, 11) p=0.836§	+7 (-9, 24) p=0.400§	+5 (-11, 20) p=0.576§	0 (-13, 13) p>0.999§	-2 (-16, 12) p=0.808§	+8 (-9, 25) p=0.369§	+5 (-10, 20) p=0.521§	0 (-16, 16) p=0.484§	-2 (-17, 13) p=0.788§	+10 (-6, 26) p=0.241§	+6 (-6, 18) p=0.320§	0 (-18, 18) p>0.999§

Within column p-values: [‡]Xpert vs. Ultra within an analysis (non-head-to-head or head-to-head) in patients of the same HIV status (overall, negative, or positive), [§]Study Ultra results (Supplementary Table 2) vs. study Ultra results excluding trace results within an analysis (non-head-to-head or head-to-head) in patients using different reference standards (MRS, eMRS, or CRS).
Within row p-values: [*]MRS vs. eMRS, [±]eMRS vs. CRS, [¥]MRS vs. CRS within an analysis (non-head-to-head or head-to-head).

[ϖ]Although Xpert data are already shown in Supplementary Table 2, small differences in the number of samples included occur in the head-to-head comparison. For the non-head-to-head comparison the Xpert data are identical to that in Supplementary Table 2 but are included here for readability.

[ϕ]This comparison is Ultra with traces excluded vs. Ultra with traces included and considered positive.

Abbreviations: CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 6: Non-head-to-head and head-to-head diagnostic accuracy analyses of Xpert and Ultra using a MRS, eMRS and CRS for the detection of Mycobacterium tuberculosis complex DNA, reclassifying trace positive Ultra results as negative. Routine Xpert results were compared to study Ultra results. Ultra has similar diagnostic accuracy compared to Xpert when trace positive results are reclassified as negative. Study Ultra results with trace reclassified had increased sensitivity and decreased specificity compared to normal study Ultra results. Similar trends are seen across reference standards. Data are %, 95% CI, and n/N.

	Non-head-to-head											
	MRS				**eMRS**				**CRS**			
	n=96				**n=97**				**n=97**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert$^{\omega}$	73 (58, 85) 35/48	92 (80, 98) 44/48	90 (76, 97) 35/39	77 (64, 87) 44/57	69 (55, 81) 36/52 p=0.685*	91 (79, 98) 41/45 p=0.924*	90 (76, 97) 36/40 p=0.510*	72 (58, 83) 41/57 p=0.519*	65 (52, 77) 39/60 p=0.635$^{\pm}$ p=0.379$^{¥}$	97 (86, 100) 36/37 p=0.244$^{\pm}$ p=0.274$^{¥}$	98 (87, 100) 39/40 p=0.166$^{\pm}$ p=0.157$^{¥}$	63 (49, 76) 36/57 p=0.317$^{\pm}$ p=0.102$^{¥}$
	n=130				**n=131**				**n=131**			
Ultra with trace reclassified	75 (62, 85) 45/60 p=0.806‡	87 (77, 94) 61/70 p=0.441‡	83 (71, 92) 45/54 p=0.379‡	80 (70, 89) 61/76 p=0.667‡	73 (61, 84) 47/64 p=0.618‡ p=0.843*	88 (78, 95) 59/67 p=0.609‡ p=0.871*	85 (73, 94) 47/55 p=0.510‡ p=0.760*	78 (67, 86) 59/76 p=0.451‡ p=0.691*	67 (55, 77) 51/76 p=0.797‡ p=0.415$^{\pm}$ p=0.316$^{¥}$	93 (82, 98) 51/55 p=0.343‡ p=0.389$^{\pm}$ p=0.310$^{¥}$	93 (82, 98) 51/55 p=0.304‡ p=0.221$^{\pm}$ p=0.130$^{¥}$	67 (55, 77) 51/76 p=0.636‡ p=0.147$^{\pm}$ p=0.066$^{¥}$
Δ Trace reclassified$^{\phi}$	-10 (-19, -1) **p=0.014**§	+18 (8, 29) **p<0.001**§	+13 (-1, 28) p=0.081§	-4 (-17, 9) p=0.558§	+10 (-18, 1) **p=0.014**§	+19 (8, 30) **p<0.001**§	+13 (-0.03, 28) p=0.063§	-3 (-17, 11) p=0.667§	-9 (-17, 1) **p=0.008**§	+22 (9, 35) **p=0.001**§	+15 (3, 26) **p=0.026**§	+1 (-17, 14) p=0.873§
	Head-to-head											
	n=92				**n=92**				**n=92**			
Xpert$^{\omega}$	72 (57, 84) 33/46	93 (82, 99) 43/46	92 (78, 98) 33/36	77 (64, 87) 43/56	67 (52, 80) 33/49 p=0.642*	93 (81, 99) 40/43 p=0.932*	92 (78, 98) 33/36 p>0.999*	71 (58, 83) 40/56 p=0.518*	64 (50, 76) 35/55 p=0.691$^{\pm}$ p=0.387$^{¥}$	97 (86, 100) 36/37 p=0.382$^{\pm}$ p=0.419$^{¥}$	97 (85, 100) 35/36 p=0.304$^{\pm}$ p=0.303$^{¥}$	64 (50, 77) 36/56 p=0.418$^{\pm}$ p=0.147$^{¥}$
Ultra with trace reclassified	78 (64, 89) 36/46 p=0.470‡	85 (71, 94) 39/46 p=0.180‡	84 (69, 93) 36/43 p=0.290‡	80 (66, 90) 39/49 p=0.729‡	76 (61, 87) 37/49 p=0.371‡ p=0.751*	86 (72, 95) 37/43 p=0.291‡ p=0.763*	86 (72, 95) 37/43 p=0.434‡ p=0.763*	76 (61, 87) 37/49 p=0.637‡ p=0.498*	73 (59, 84) 40/55 p=0.306‡ p=0.747$^{\pm}$ p=0.521$^{¥}$	92 (78, 98) 34/37 p=0.304‡ p=0.409$^{\pm}$ p=0.323$^{¥}$	93 (81, 99) 40/43 p=0.397‡ p=0.291$^{\pm}$ p=0.178$^{¥}$	69 (55, 82) 34/49 p=0.580‡ p=0.489$^{\pm}$ p=0.247$^{¥}$
Δ Trace reclassified$^{\phi}$	-13 (-25, 1) **p=0.014**§	+9 (-2, 19) **p=0.046**§	+5 (-11, 20) p=0.576§	-10 (-25, 5) p=0.196§	-12 (-23, -1) **p=0.014**§	+9 (-2, 20) **p=0.046**§	+5 (-10, 20) p=0.521§	-9 (-26, 7) p=0.293§	-11 (-21, -1) **p=0.014**§	+11 (-2, 24) **p=0.046**§	+6 (-6, 18) p=0.320§	-8 (-26, 11) p=0.430§

Within column p-values: [‡]Xpert vs. Ultra within an analysis (non-head-to-head or head-to-head) in patients of the same HIV status (overall, negative, or positive), [§]Study Ultra results (Supplementary Table 2) vs. study Ultra results excluding trace results within an analysis (non-head-to-head or head-to-head) in patients using different reference standards (MRS, eMRS, or CRS).

Within row p-values: [*]MRS vs. eMRS, [±]eMRS vs. CRS, [¥]MRS vs. CRS within an analysis (non-head-to-head or head-to-head).

[ø]Although Xpert data are already shown in Supplementary Table 2, small differences in the number of samples included occur in the head-to-head comparison. For the non-head-to-head comparison the Xpert data are identical to that in Supplementary Table 2 but are included here for readability.

[ϕ]This comparison is Ultra with traces reclassified as negative vs. Ultra with traces considered positive.

Abbreviations: CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Appendix II

Diagnostic accuracy of Xpert MTB/RIF Ultra on pericardial fluid and urine for tuberculosis pericarditis diagnosis in an HIV-endemic setting

Chapter 3

(Supplementary material)

Table of Contents

Methods

Definitions

Actionable results

Xpert and Ultra: For TB, positive or negative. For rifampicin, TB-positive and resistant or susceptible.

Culture: Positive or negative for MTBC.

PF descriptions

Blood stained: Filled with blood.

Pyopericardium: Accumulation of pus.

Serous: Clear to pale yellow.

Serous-sanguineous: Clear with a small amount of blood.

Information of the programmatic diagnostic algorithm

This differed based on if patients had PF or biopsies collected and tested by the National Health Laboratory Service (NHLS) within three months prior to recruitment. If this was met and previous samplings were Xpert- or Ultra-positive, 5-7.5 ml of the presently collected PF and/or biopsy was NALC-NaOH decontaminated, neutralised, centrifuged, and resuspended in phosphate buffer, with 500 µl inoculated into culture (**Figure 1)**. If these previous samplings were Xpert- or Ultra-negative or previous samples were not collected, Xpert or Ultra were done together with culture on the presently collected specimens.

Results

Comparison of study unconcentrated Ultra true-positives vs. false-positives results (per MRS)

True-positives vs. false-positives were more likely to have a pleural effusion visible on chest X-ray [73% (35/48) vs. 47% (14/30), p=0.020], have serosanguineous PF compared to all other types [29% (14/48) vs. 10% (3/30), p=0.046], experienced pericardial tamponade [40% (17/42) vs. 14% (4/28), p=0.019], and higher ADA [62.10 (IQR: 50.50-82.80) vs. 20.90 (4.70-47.90); p<0.001] and LDH [1073 (554-1948) vs. 539 (191-1146); p=0.009] levels.

uIFN-γ on PF

Cut-points under different scenarios: Overall area under the ROC curve (AUROC) was 0.76 [95% confidence interval (CI): 0.69, 084] (**Figure 4A**). Rule-out cut-point sensitivity and specificity is in the main text. At a rule-in cut-point (>4277 pg/ml), specificity was 95% (95% CI: 90, 98) and sensitivity 8% (3, 17). At a cut-point corresponding to Youden's index (>125.8 pg/ml) sensitivity was 92% (83, 97) and specificity 64% (55, 73).

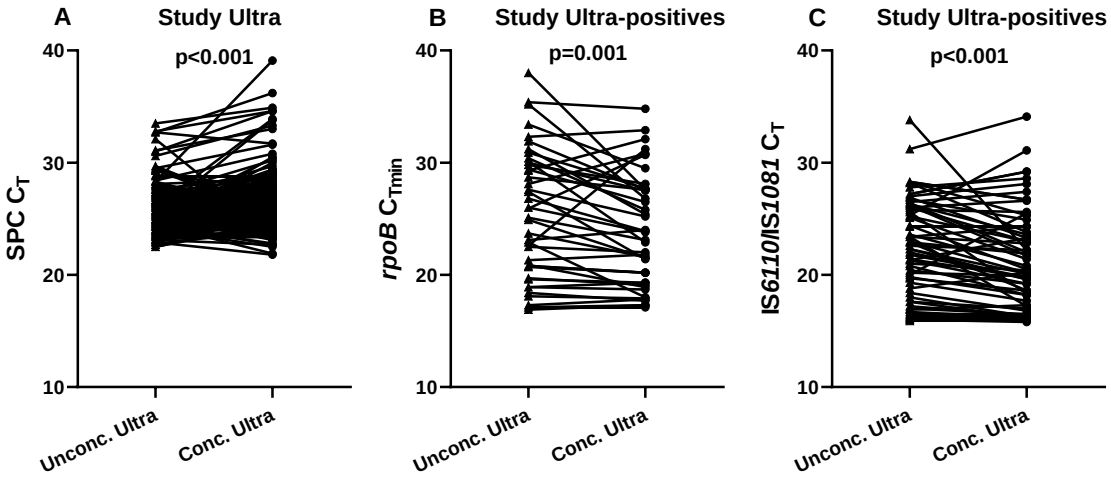
*HIV-positive*s: AUROC was 0.64 (0.52, 0.72) (**Figure 4B)**. At a rule-out cut-point (>1.2 pg/ml), sensitivity was 95% (86, 99) and specificity 24% (13, 39). At a rule-in cut-point (>4572 pg/ml), sensitivity was 7% (2, 17) and specificity was 95% (84, 99). At the Youden's index cut-point (>303.2 pg/ml), sensitivity was 88% (77, 95) and specificity 49% (34, 63).

HIV-negatives: AUROC was 0.87 (0.79, 0.96) (**Figure 4C**). At a rule-out cut-point (>125.8 pg/ml), sensitivity was 94% (75, 100) and specificity 78% (66, 87). At a rule-in cut-point (>3307.4 pg/ml), sensitivity was 24% (9, 46) and specificity was 96% (88, 99). At a cut-point of >125.8 pg/ml (Youden's index), sensitivity was 94% (75, 100) and specificity 78% (66, 87).

Study vs. programmatic Ultra PF results

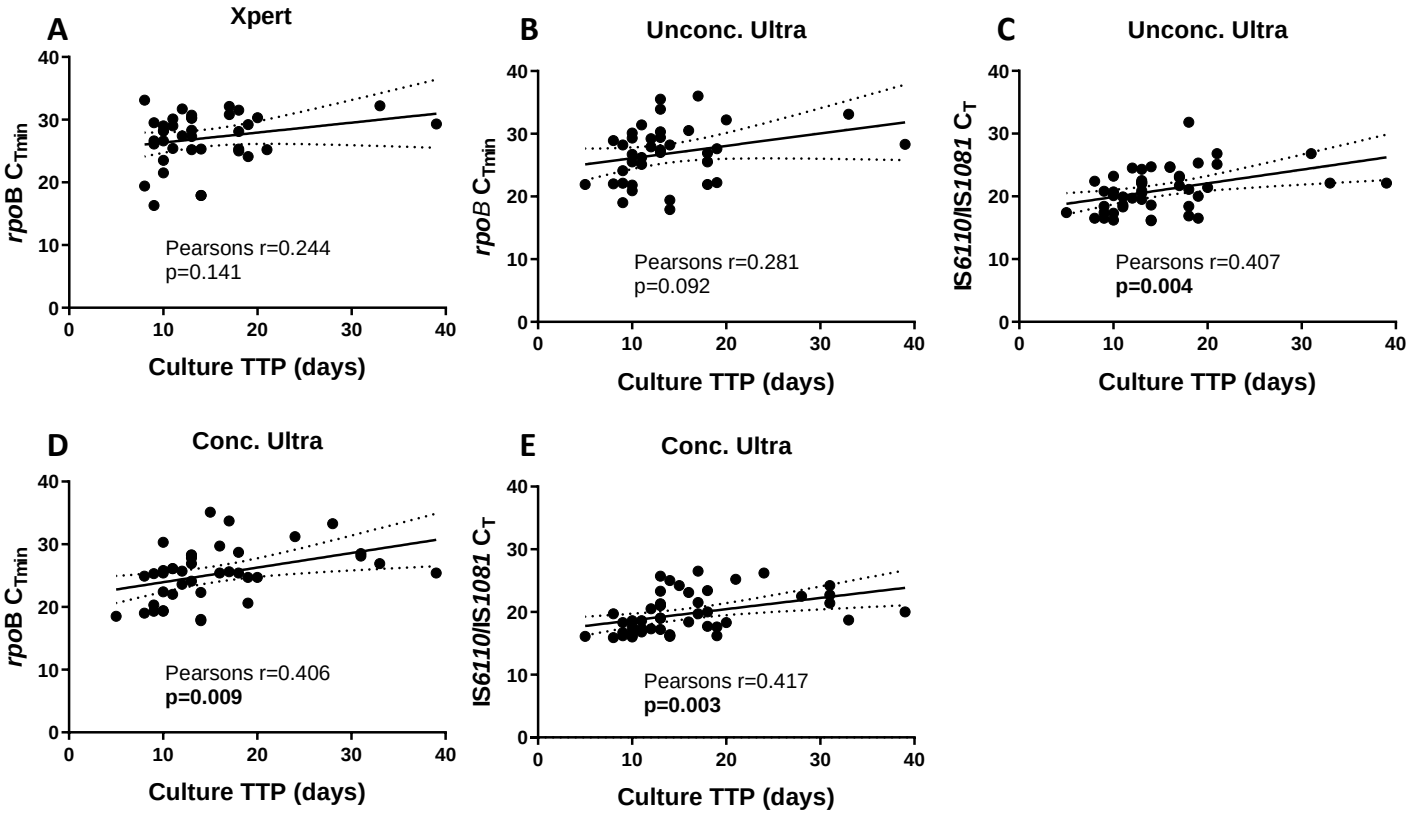
In patients who received two concentrated Ultras (study and programmatic), positivity rates were similar [41% (21/51) vs. 43% (22/51); p=0.655] (**Supplementary Table 10**).

Supplementary Figure 1: Spaghetti plots showing PCR quantitative information (C_T) for different tests to measure the effect of PF concentration. PCR inhibition increased with concentration (higher SPC C_T indicates more inhibition) (**A**) whereas mycobacterial load measured using *rpoB* and IS*6110*/IS*1081* probes (**B, C**) increased (lower values indicate higher template DNA concentrations). Abbreviations: Conc., concentrated; PF, pericardial fluid; C_T, cycle threshold value; C_{Tmin}, minimum C_T; SPC; sample processing control; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated.



*Two zero SPC values excluded from B

Supplementary Figure 2: Xpert and Ultra quantitative information versus MGIT960 TTP. (**A**) Shows Xpert whereas **(B-E)** shows unconc. and conc. Ultra results for each probe. A significant positive linear correlation between IS*6110*/IS*1081* C_T and TTP was observed regardless of PF concentration. For Ultra *rpo*B C_{Tmin}, this was only observed after concentration. Xpert included programmatic and study results (and hence unconc. and conc. results) and, for Ultra, only study results are shown (as routine conc. Ultras showed no difference in positivity compared study conc. Ultras). Dashed lines are 95% confidence intervals. Abbreviations: Conc., concentrated; MGIT, mycobacteria growth indicator tube; PF, pericardial fluid; TTP, culture time-to-positivity; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.



Supplementary Table 1: Reference standard definitions.

	MRS*	eMRS†	CRS‡
	Site of disease fluid		
MGIT960 culture on PF	✓	✓	✓
MGIT960 culture on pericardial biopsy	✓	✓	✓
	Non-site-of disease fluid		
Smear microscopy	×	✓	✓
Xpert	×	✓	✓
Ultra	×	✓	✓
MGIT960	×	✓	✓
	Alternate diagnosis and treatment information		
Alternate diagnosis	×	×	✓
TB treatment initiated	×	×	✓
	Case definitions		
Reference standard positive (Definite-TB)	Any MRS test positive	Any eMRS test positive	Any eMRS test positive or TB treatment was initiated and no alternate diagnosis
Reference standard negative (Non-TB)	No MRS test positive	No eMRS test positive	No eMRS test positive and patient not initiated on treatment with and without an alternate diagnosis
Probable TB	N/A	N/A	N/A
Unclassifiable	No positive MRS test and site-of-disease fluid culture contaminated or not done	No positive eMRS test and site-of-disease fluid culture contaminated or not done	No positive eMRS test and site-of-disease fluid culture contaminated or not done, or treatment not initiated, or no alternate diagnosis available

Abbreviations: CRS, composite reference standard; eMRS, extended reference standard; MGIT960 culture Mycobacteria Growth Indicator Tube 960; MRS, microbiological reference standard; PF, pericardial fluid; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 2: Xpert (study and programmatic) and Ultra (study) RIF results from PF compared to MTBDR*plus* results from culture-positive isolates. Ultra detected all RIF-resistant results detected by MTBDR*plus* while Xpert missed one case. Data are % and n/N.

MTBDR*plus* result	Xpert-positive*	Unconc. Ultra-positive	Conc. Ultra-positive
Resistant (2/41)	**RIF-resistant**		
	50 (1/2)	100 (2/2)	100 (2/2)
	RIF-susceptible		
	50 (1/2)	0 (0/2)	0 (0/2)
	RIF-indeterminate		
	0 (0/2)	0 (0/2)	0 (0/2)
Susceptible (39/41)	**RIF-resistant**		
	4 (1/25)	0 (0/32)	0 (0/31)
	RIF-susceptible		
	88 (22/25)	72 (23/32)	81 (25/31)
	RIF-indeterminate		
	8 (2/25)	28 (9/32) (all trace-positive)	19 (6/31) (all trace-positive)

*Xpert included study unconcentrated and programmatic concentrated results.

Missing data: Xpert not done, 2; Non-actionable study concentrated Ultras, 1.

Abbreviations: Conc., concentrated; MTBDR*plus*, GenoType MTBDR*plus*; PF, pericardial fluid; RIF, rifampicin; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Supplementary Table 3: Sensitivity and specificity for different cut-off points of uIFN-γ, ADA and albumin stratified by HIV status. uIFN-γ had higher overall sensitivity and specificity compared to ADA and albumin, and performed better in PLHIV. Sensitivity is low in all rule-in scenarios for each marker (including stratified by HIV) and for rule-out scenarios, uIFN-γ has the best sensitivity and specificity (with ADA coming close in people without HIV). Data are % and 95% CI.

	uIFN-γ								
	Overall			**HIV-negative**			**HIV-positive**		
	Rule-in (>4277 pg/mL)	**Rule-out (>5.1 pg/mL)**	**Youden's index (>125.8 pg/mL)**	**Rule-in (>3307.4 pg/mL)**	**Rule-out (>125.8 pg/mL)**	**Youden's index (>125.8 pg/mL)**	**Rule-in (>4572 pg/mL)**	**Rule-out (>1.2 pg/mL)**	**Youden's index (>303.2 pg/mL)**
Sensitivity	8 (3, 17)	95 (88, 99)	92 (83, 97)	21 (8, 42)	95 (77, 100)	95 (77, 100)	7 (2, 17)	95 (85, 99)	88 (77, 95)
Specificity	95 (90, 98)	53 (44, 62)	64 (55, 73)	96 (88, 99)	78 (66, 87)	78 (66, 87)	95 (84, 99)	24 (13, 39)	49 (34, 63)
	ADA								
	Rule-in (>145.5 U/L)	**Rule-out (>0 U/L)**	**Youden's index (>38.6 U/L)**	**Rule-in (>166.3 U/L)**	**Rule-out (>33.0 U/L)**	**Youden's index (>33.0 U/L)**	**Rule-in (>133.5 U/L)**	**Rule-out (>0 U/L)**	**Youden's index (>38.6 U/L)**
Sensitivity	0 (0, 5)	93 (85, 98)	86 (77, 93)	0 (0, 15)	95 (77, 100)	95 (77, 100)	0 (0, 7)	93 (82, 98)	85 (73, 93)
Specificity	95 (89, 98)	9 (4, 15)	60 (50, 69)	96 (88, 99)	65 (53, 77)	65 (53, 77)	94 (82, 99)	9 (3, 22)	46 (31, 61)
	Albumin								
	Rule-in (<2 g/L)	**Rule-out (<29 g/L)**	**Youden's index (<25 g/L)**	**Rule-in (<5 g/L)**	**Rule-out (<31 g/L)**	**Youden's index (<26 g/L)**	**Rule-in (<2 g/L)**	**Rule-out (<28 g/L)**	**Youden's index (<21 g/L)**
Sensitivity	0 (0, 5)	97 (90, 99)	78 (67, 86)	0 (0, 15)	95 (78, 100)	84 (64, 96)	0 (0, 7)	95 (85, 99)	58 (43, 71)
Specificity	100 (96, 100)	23 (16, 32)	46 (37, 56)	100 (94, 100)	22 (13, 35)	53 (40, 65)	100 (91, 100)	6 (1, 18)	61 (45, 75)

Abbreviations: ADA, adenosine deaminase; uIFN-γ, unstimulated interferon-γ.

Supplementary Table 4: All initial non-actionable results for study Xpert (unconcentrated; information from programmatic testing unavailable) and study Ultra (unconcentrated, concentrated). If a specimen had sufficient volume, the test was repeated once. Majority of non-actionable results resolved after re-testing.

Test	Patient ID	First result	Error code	Second result	Error code/reason test was not repeated
Xpert (study unconc. only)	PCTB034	Error	5007	Positive	-
	PCTB037	No result	2005	Negative	-
	PCTB044	No result	5006	Error	5006
	PCTB059	Error	5017	Negative	-
	PCTB067	Error	5007	Error	5007
	PCTB069	Error	5011	Negative	-
	PCTB075	No result	2005	Negative	-
	PCTB109	No result	2037	Negative	-
	PCTB110	Error	5011	Positive	-
	PCTB111	Error	5007	Positive	-
80% (8/10) non-actionable results resolved upon repeat testing (3 Xpert-positive, 5 Xpert-negative)					
Unconc. Ultra	PCTB020	Error	5007	Positive	-
	PCTB038	Error	5006	Negative	-
	PCTB039	Error	5006	Negative	-
	PCTB101	Error	5007	Positive	-
100% (4/4) non-actionable results resolved upon repeat testing (2 Ultra-positive, 2 Ultra-negative)					
Conc. Ultra	PCTB038	Error	5006	Positive	-
	PCTB058	Error	5007	Negative	-
	PCTB063	Error	5007	Negative	-
	PCTB069	Error	5007	Error	5007
	PCTB071	Error	5007	Error	5007
	PCTB073	Error	2008	Positive	-
	PCTB074	Error	2008	Positive	-
	PCTB106	Error	2008	Negative	-

	PCTB108	Error	2008	Error	2008
	PCTB109	Error	5007	Negative	-
	PCTB113	Error	2008	Error	2008
	PCTB116	Error	2008	Not done	No more fluid
	PCTB120	Error	5007	Not done	No more fluid
	PCTB124	Error	2008	Error	2008
	PCTB126	Error	2008	Positive	-
	PCTB145	Error	5011	Positive	-
	PCTB149	Error	5007	Negative	-
	PCTB155	Error	5007	Negative	-
69% (11/16) non-actionable results resolved upon repeat testing (5 Ultra-positive, 6 Ultra-negative)					

Error code definitions:
2005: Motion of the syringe drive was not detected.
2037: The cartridge integrity test failed.
2005: Motion of the syringe drive was not detected.
2008: Syringe pressure reading the protocol limit.
5006: Probe D check failed.
5011: Signal loss detected in the amplification curve for analyte SPC.
5017: SPC probe check failed.

Abbreviations: Conc., concentrated; unconc., unconcentrated; MRS, microbiological reference standard; uIFN-γ, unstimulated interferon-γ; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Supplementary Table 5: Comparison of actionable and non-actionable study unconc. and conc. Ultra results. Conc. Ultra non-actionables had lower uIFN-γ and protein concentrations. Data are n/N (%) or median (IQR).

	Unconc.		Conc.	
	Actionable	**Non-actionable**	**Actionable**	**Non-actionable**
PF characteristics				
Bloody	3/4 (75)	67/151 (44) p=0.244	62/137 (45)	8/18 (44) p=0.948
Chylous	0/4 (0)	1/151 (1) p=0.870	1/137 (1)	0/18 (0) p=0.716
Purulent	0/4 (0)	6/151 (4) p=0.684	6/137 (4)	0/18 (0) p=0.365
Serous	1/4 (25)	44/151 (29) p=0.857	38/137 (28)	7/18 (39) p=0.327
Serous-sanguineous	0/4 (0)	33/151 (22) p=0.292	30/137 (22)	3/18 (17) p=0.610
Clinical characteristics				
HIV-positive	3/4 (75)	80/149 (54) p=0.399	77/137 (56)	6/16 (38) p=0.155
TB treatment				
Previous TB	2/3 (66)	32/149 (21) p=0.063	30/134 (22)	4/18 (22) p=0.987
Fluid biomarkers				
uIFN-γ (pg/ml)	701 (1-2460)	1310 (197-2179) p=0.892	906 (3-2576)	0 (0-649) **p=0.002**
ADA (U/L)	45.9 (21.7-64.9)	27.2 (2.9-61.4) p=0.368	46.7 (19.2-65.2)	31.5 (22.2-62.9) p=0.435
Albumin (g/L)	21.5 (16.0-26.0)	18.5 (16.0-21.8) p=0.364	22 (17-26)	18 (7-23) p=0.077
Protein (g/L)	57.5 (49.0-64.3)	53 (33.0-64.8) p=0.557	59 (51-66)	50 (42-62) **p=0.016**
LDH (U/L)	655.5 (428.8-1446)	864.5 (553.8-1808) p=0.680	645 (440-1380)	929 (173-2260) p=0.625

Missing data: HIV, 2; previous TB, 3; uIFN-γ, 3; ADA, 9; albumin, 9, protein, 9; LDH, 9.

Abbreviations: ADA, Adenosine deaminase; conc., concentrated; LDH, lactate dehydrogenase; unconc., unconcentrated; uIFN-γ, unstimulated interferon-γ; unconc., unconcentrated.

Supplementary Table 6: The change in Ultra (unconc. and conc.) diagnostic accuracy when Ultra traces were either excluded or reclassified as negative when compared to the MRS. Unconc. Ultra had increased specificity when traces were excluded and both decreased sensitivity and increased specificity when traces were reclassified. Conc. Ultra showed no change in diagnostic accuracy when traces were excluded or reclassified. Conclusions were similar to non-head-to-head analyses (**Supplementary Table 6**). Data are %, 95% CI, and n/N.

	All patients				HIV-negative				HIV-positive			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
	n=135				**n=62/135 (46)**				**n=73/135 (54)**			
Δ unconc. Ultra (study) traces excluded	-3% (-19, 13) p=0.697	+18% (5, 31) **p=0.010**	+17% (0.1, 33) p=0.064	+1% (-12, 14) p=0.872	-4% (-38, 30) p=0.822	+26% (9, 43) **p=0.009**	+30% (-2, 62) p=0.093	+3% (-14, 19) p=0.756	-3% (-22, 14) p=0.698	+9% (-11, 29) p=0.376	+5% (-13, 24) p=0.580	-1% (-20, 19) p=0.951
Δ conc. Ultra (study) traces excluded	+1% (-13, 15) p=0.934	-0.2% (-13, 12) p=0.970	+0.03% (-16, 16) p=0.997	+0.3% (-10, 11) p=0.958	+1% (-27, 29) p=0.945	-1% (-17, 13) p=0.852	-2% (-33, 29) p=0.886	+0.3% (-11, 12) p=0.959	+0.3% (-16, 16) p=0.975	+2% (-19, 23) p=0.824	+1% (-18, 19) p=0.929	+1% (-17, 20) p=0.890
Δ unconc. Ultra (study) traces reclassified	-20% (-37, 3) **p=0.022**	+21% (9, 33) **p=0.001**	+17% (0.1, 33) p=0.064	-7% (-19, 6) p=0.300	-24% (-56, 9) p=0.163	+29% (13, 45) **p=0.001**	+29% (-2, 62) p=0.093	-3% (-19, 13) p=0.720	-18% (-37, 1) p=0.064	+11% (-7, 30) p=0.232	+5% (-13, 24) p=0.580	-11% (-31, 8) p=0.258
Δ conc. Ultra (study) traces reclassified	-15% (-32, 8) p=0.065	+3% (-9, 14) p=0.668	-1% (-16, 15) p=0.942	-8% (-19, 3) p=0.150	-24% (-56, 9) p=0.163	+29% (13, 45) **p=0.001**	+29% (-2, 62) p=0.093	-3% (-19, 13) p=0.720	-18% (-37, 1) p=0.064	+11% (-7, 30) p=0.232	+5% (-13, 24) p=0.580	-11% (-31, 8) p=0.258

Abbreviations: Conc., concentrated; CI, confidence interval; MRS, microbiological reference standard; NPV, negative predictive value; PF, pericardial fluid; PPV, positive predictive value; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated.

Supplementary Table 7: Non-head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on PF stratified by HIV status. Conclusions were similar to head-to-head analyses (**Table 2**). Data are %, 95% CI, and n/N.

	All patients				**HIV-negative**				**HIV-positive**			
	n=147				**n=70/147 (48)**				**n=77/147 (52)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	66 (52, 78) 38/58	93 (86, 97) 83/89	86 (73, 95) 38/44	81 (72, 88) 83/103	67 (41, 87) 12/18	94 (84, 99) 49/52	80 (52, 96) 12/15	89 (78, 96) 49/55	65 (48, 79) 26/40 p=0.902*	92 (78, 98) 34/37 p=0.665*	90 (73, 98) 26/29 p=0.376*	71 (56, 83) 34/48 **p=0.019***
	n=150				**n=71/150 (47)**				**n=79/150 (53)**			
Unconc. Ultra (study)	80 (68, 89) 48/60 p=0.077‡	67 (56, 76) 60/90 **p<0.001‡**	62 (50, 72) 48/78 **p=0.004‡**	83 (73, 91) 60/72 p=0.643‡	68 (43, 87) 13/19 p=0.909‡	60 (45, 73) 31/52 **p=0.001‡**	38 (22, 56) 13/34 **p=0.007‡**	84 (68, 94) 31/37 p=0.459‡	85 (71, 94) 35/41 **p=0.034‡** p=0.127*	76 (60, 89) 29/38 p=0.066‡ p=0.097*	80 (65, 90) 35/44 p=0.254‡ **p<0.001***	83 (66, 93) 29/35 p=0.206‡ p=0.916*
	n=140				**n=64/140 (46)**				**n=76/140 (54)**			
Conc. Ultra (study)	86 (74, 94) 49/57 p=0.392±	83 (73, 90) 69/83 **p=0.013±**	78 (66, 87) 49/63 **p=0.039±**	90 (81, 95) 69/77 p=0.261±	83 (59, 96) 15/18 p=0.291±	89 (76, 96) 41/46 **p=0.001±**	75 (51, 91) 15/20 **p=0.009±**	93 (81, 99) 41/44 p=0.180±	87 (73, 96) 34/39 p=0.814± p=0.698*	76 (59, 88) 28/37 p=0.948± p=0.102*	79 (64, 90) 34/43 p=0.956± p=0.718*	85 (68, 95) 28/33 p=0.824± p=0.236*
	n=147				**n=69/147 (47)**				**n=78/147 (53)**			
uIFN-γ	95 (86, 99) 57/60 **p=0.013¥**	53 (42, 64) 46/87 **p=0.061¥**	58 (48, 68) 57/98 p=0.650¥	94 (83, 99) 46/49 p=0.084¥	100 (82, 100) 19/19 **p=0.008¥**	64 (49, 77) 32/50 p=0.649¥	51 (34, 68) 19/37 p=0.267¥	100 (89, 100) 32/32 **p=0.017¥**	93 (80, 98) 38/41 p=0.289¥ p=0.226*	38 (22, 55) 14/37 **p=0.001¥** **p=0.016***	62 (49, 74) 38/61 p=0.058¥ p=0.287*	82 (57, 96) 14/17 p=0.964¥ **p=0.014***
Δ unconc. Ultra (study) traces excluded	-4% (-21, 12) p=0.582	+18% (5, 31) **p=0.008**	+17% (0.5, 33) p=0.055	0% (-12, 12) p=0.003	-8% (-41, 24) p=0.610	+29% (12, 46) **p=0.003**	+31% (1, 61) p=0.057	0% (-17, 17) p>0.999	-3% (-20, 14) p=0.723	+9% (-9, 27) p=0.337	+5% (-12, 22) p=0.550	0% (-18, 18) p>0.999
Δ conc. Ultra (study) traces excluded	-2% (-16, 11) p=0.743	+2% (-9, 13) p=0.719	-0.4% (-17, 15) p=0.957	0% (-10, 10) p>0.999	-5% (-32, 23) p=0.732	+2% (-10, 14) p=0.752	-2% (-31, 28) p=0.911	0% (-10, 10) p>0.999	-1% (-17, 14) p=0.854	+2% (-17, 21) p=0.832	-0.1% (-18, 18) p=0.989	0% (-17, 17) p>0.999
Δ unconc. Ultra (study) traces reclassified	-18% (-35, 2) **p=0.031**	+21% (9, 33) **p=0.001**	+17% (0.5, 33) p=0.055	-5% (-17, 7) p=0.387	-21% (-52, 10) p=0.189	+32% (18, 49) **p<0.001**	+31% (1, 61) p=0.057	-1% (-16, 14) p=0.896	-17% (-35, 1) p=0.067	+11% (-7, 28) p=0.237	+5% (-12, 22) p=0.550	-11% (-29, 7) p=0.242

Δ conc. Ultra (study) traces reclassified	-14% (-29, 1) p=0.066	+3% (-9, 13) p=0.669	-0.4% (-17, 15) p=0.957	-8% (-19, 3) p=0.148	-22% (-51, 6) p=0.137	+2% (-10, 14) p=0.726	-2% (-31, 28) p=0.911	-7% (-20, 5) p=0.246	-10% (-27, 7) p=0.238	+3% (-16, 22) p=0.782	-0.1% (-18, 18) p=0.989	-9% (-27, 10) p=0.367

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ±unconc. Ultra (study) vs. conc. Ultra (study), ¥unconc. Ultra (study) vs. uIFN-γ in patients of the same HIV status (overall, negative, positive).
Within row p-values: *HIV-negative vs. HIV-positive.

Abbreviations: Conc., concentrated; CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PF, pericardial fluid; PPV, positive predictive value; uIFN-γ, unstimulated interferon-γ; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Supplementary Table 8: Head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on PF compared to a MRS, eMRS and CRS for the detection of Mycobacterium tuberculosis complex DNA. Compared to the MRS, the CRS had decreased sensitivity and similar specificity for Xpert and Ultra (unconc. and conc.). uIFN-γ had similar sensitivity and increased specificity compared to the MRS. The relative performance of eMRS compared to the MRS was the same **(Table 2)**. Data are %, 95% CI, and n/N.

	MRS				**eMRS**				**CRS**			
	n=135				**n=135**				**n=135**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic unconc. and conc. pooled)	64 (50, 76) 35/55	93 (84, 97) 74/80	85 (71, 94) 35/41	79 (69, 86) 74/94	56 (43, 68) 38/68 p=0.384*	96 (87, 99) 64/67 p=0.447*	93 (80, 98) 38/41 p=0.289*	68 (58, 77) 64/94 p=0.099*	39 (30, 50) 39/99 **p=0.036§** **p=0.004†**	94 (81, 99) 34/36 p=0.808§ p=0.702†	95 (83, 99) 39/41 p=0.644§ p=0.137†	36 (27, 47) 34/94 **p<0.001§** **p<0.001†**
Unconc. Ultra (study)	80 (67, 90) 44/55 p=0.057‡	69 (57, 79) 55/80 **p<0.001‡**	64 (51, 75) 44/69 **p=0.015‡**	83 (72, 91) 55/66 p=0.478‡	72 (60, 82) 49/68 **p=0.049‡** p=0.308*	70 (58, 81) 47/67 **p<0.001‡** p=0.855*	71 (59, 81) 49/69 **p=0.007‡** p=0.364*	71 (59, 82) 47/66 p=0.673‡ p=0.097*	58 (47, 67) 57/99 **p=0.011‡** p=0.056§ **p=0.005†**	67 (49, 81) 24/36 **p=0.003‡** p=0.716§ p=0.824†	83 (72, 91) 57/69 p=0.057‡ p=0.107§ **p=0.013†**	36 (25, 49) 24/66 p=0.980‡ **p<0.001§** **p<0.001†**
Conc. Ultra (study)	85 (73, 94) 47/55 p=0.449±	83 (72, 90) 66/80 **p=0.043±**	77 (65, 87) 47/61 p=0.099±	89 (80, 95) 66/74 p=0.313±	78 (66, 87) 53/68 p=0.428± p=0.288*	88 (78, 95) 59/67 **p=0.011±** p=0.347*	87 (76, 94) 53/61 **p=0.028±** p=0.158*	80 (69, 88) 59/74 p=0.241± p=0.112*	60 (49, 69) 59/99 p=0.773± **p=0.013§** **p=0.001†**	94 (81, 99) 34/36 **p=0.003±** p=0.297§ p=0.084†	97 (89, 100) 59/61 **p=0.010±** **p=0.048§** **p=0.001†**	46 (34, 58) 34/74 p=0.251± **p<0.001§** **p<0.001†**
uIFN-γ (rule-out cut-off 5.1 pg/ml)	95 (85, 99) 52/55 **p=0.022¥**	50 (39, 61) 40/80 **p=0.016¥**	57 (46, 67) 52/92 p=0.354¥	93 (81, 99) 40/43 p=0.140¥	96 (88, 99) 65/68 **p<0.001¥** p=0.790*	60 (47, 72) 40/67 p=0.205¥ p=0.240*	71 (60, 80) 65/92 p=0.960¥ **p=0.004***	93 (81, 99) 40/43 **p=0.006¥** p>0.999*	86 (77, 92) 85/99 **p<0.001¥** **p=0.041§** p=0.099†	81 (64, 92) 29/36 p=0.181¥ **p=0.032§** **p=0.002†**	92 (85, 97) 85/92 p=0.057¥ **p<0.001§** **p<0.001†**	67 (51, 81) 29/43 **p=0.002¥** **p=0.003§** **p=0.003†**

Missing data: Xpert non-actionable, n=2; Xpert not done, n=1; study concentrated Ultra non-actionable, n=7; study concentrated Ultra not done, n=3; uIFN-γ not done, n=3.

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ±unconc. Ultra (study) vs. conc. Ultra (study), ¥unconc. Ultra (study) vs. uIFN-γ in patients using the same reference standard (MRS, eMRS, CRS).

Within row p-values: *MRS vs. eMRS, §eMRS vs. CRS, †MRS vs. CRS.

Abbreviations: conc., concentrated; CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PF, pericardial fluid; PPV, positive predictive value; uIFN-γ, unstimulated interferon gamma; Ultra; Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Supplementary Table 9: Non-head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on PF using a MRS, eMRS and CRS for the detection of Mycobacterium tuberculosis complex DNA. Conclusions were like head-to-head diagnostic accuracy analyses (**Supplementary Table 8**). Data are %, 95% CI, and n/N.

	MRS				eMRS				CRS			
	n=147				n=149				n=152			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic unconc. and conc. pooled)	66 (52, 78) 38/58	93 (86, 97) 83/89	86 (73, 95) 38/44	81 (72, 88) 83/103	58 (46, 70) 42/72 p=0.403*	96 (89, 99) 74/77 p=0.419*	93 (82, 99) 42/45 p=0.276*	71 (61, 80) 74/104 p=0.113*	40 (31, 50) 43/108 **p=0.015§** **p=0.002†**	95 (85, 99) 42/44 p=0.863§ p=0.616†	96 (85, 99) 43/45 p=0.645§ p=0.130†	39 (30, 49) 42/107 **p<0.001§** **p<0.001†**
	n=150				n=152				n=155			
Unconc. Ultra (study)	80 (68, 89) 48/60 p=0.077‡	67 (56, 76) 60/90 **p<0.001‡**	62 (50, 72) 48/78 **p=0.004‡**	83 (73, 91) 60/72 p=0.643‡	72 (60, 81) 53/74 p=0.092‡ p=0.263*	67 (55, 77) 52/78 **p<0.001‡** p>0.999*	67 (56, 77) 53/79 **p=0.001‡** p=0.468*	71 (59, 81) 52/73 p=0.991‡ p=0.082*	58 (48, 67) 64/111 **p=0.008‡** p=0.054§ **p=0.003†**	64 (48, 78) 28/44 **p<0.001‡** p=0.735§ p=0.729†	80 (70, 88) 64/80 **p=0.017‡** p=0.065§ **p=0.011†**	37 (26, 49) 28/75 p=0.793‡ **p<0.001§** **p<0.001†**
	n=140				n=142				n=144			
Conc. Ultra (study)	86 (74, 94) 49/57 p=0.392±	83 (73, 90) 69/83 **p=0.013±**	78 (66, 87) 49/63 **p=0.039±**	90 (81, 95) 69/77 p=0.261±	77 (66, 87) 55/71 p=0.420± p=0.221*	89 (79, 95) 63/71 **p=0.001±** p=0.322*	87 (77, 94) 55/63 **p=0.005±** p=0.159*	80 (69, 88) 63/79 p=0.222± p=0.088*	58 (48, 68) 61/105 p=0.948± **p=0.008§** **p<0.001†**	95 (83, 99) 37/39 **p=0.001±** p=0.284§ p=0.073†	97 (89, 100) 61/63 **p=0.003±** **p=0.048§** **p=0.001†**	46 (35, 57) 37/81 p=0.291± **p<0.001§** **p<0.001†**
	n=147				n=149				n=152			
uIFN-γ (rule-out cut-off 5.1 pg/ml)	95 (86, 99) 57/60 **p=0.013¥**	53 (42, 64) 46/87 **p=0.061¥**	58 (48, 68) 57/98 p=0.650¥	94 (83, 99) 46/49 p=0.084¥	96 (89, 99) 71/74 **p<0.001¥** p=0.792*	61 (49, 72) 46/75 p=0.492¥ p=0.278*	71 (61, 80) 71/100 p=0.573¥ p=0.059*	94 (83, 99) 46/49 **p=0.002¥** p>0.999*	86 (78, 92) 94/109 **p<0.001¥** **p=0.031§** p=0.077†	84 (69, 93) 36/43 **p=0.034¥** **p=0.011§** **p=0.001†**	93 (86, 97) 94/101 **p=0.009¥** **p<0.001§** **p<0.001†**	71 (56, 83) 36/51 **p<0.001¥** **p=0.002§** **p=0.002†**

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ±unconc. Ultra (study) vs. conc. Ultra (study), ¥unconc. Ultra (study) vs. uIFN-γ in patients using the same reference standard (MRS, eMRS, CRS).

Within row p-values: *MRS vs. eMRS, §eMRS vs. CRS, †MRS vs. CRS.

Abbreviations: conc., concentrated; CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PF, pericardial fluid; PPV, positive predictive value; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF

Supplementary Table 10: Study Conc. and unconc. Ultra vs. programmatic conc. Ultra and study conc. Ultra vs. study unconc. Ultra on PF. Concordance was not observed when pairs were compared.

A		**Conc. Ultra (study)**		
		Positive	Negative	Total
Conc. Ultra (programmatic)	Positive	19	3	22
	Negative	2	27	29
	Total	21	30	51
	Δ Conc. Ultra (study) vs. conc. Ultra (programmatic)	+2% (95% CI: 9, 12) p=0.655		
B		**Unconc. Ultra (study)**		
		Positive	Negative	Positive
Conc. Ultra (programmatic)	Positive	17	6	23
	Negative	11	22	33
	Total	28	28	56
	Δ Unonc. Ultra (study) vs. conc. Ultra (programmatic)	+9% (95% confidence interval; CI: 7, 25) p=0.225		
C		**Conc. Ultra (study)**		
		Positive	Negative	Positive
Unconc. Ultra (study)	Positive	52	20	72
	Negative	11	61	72
	Total	63	81	144
	Δ Unonc. Ultra (study) vs. conc. Ultra (study)	+6% (95% confidence interval; CI: 14, 20) p=0.106		

Non-actionable rates of programmatic conc. Ultras are unknown. Five study conc. Ultras were non-actionable (1 programmatically-positive, 4 negative) and study unconc. Ultra had no non-actionable results (repeats were done with sufficient PF per patient).

Abbreviations: CI, confidence interval;conc., concentrated; PF, pericardial fluid; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated.

Supplementary Table 11: Per patient information for study unconcentrated Ultra-positive patients that were MRS-negative with Ultra semi-quantitation category, previous TB status, programmatic Xpert or Ultra results, alternate reference standard result and treatment initiation status after at least 12-weeks follow-up. Approximately 50% MRS-negative patients had TB previously, were eMRS- or CRS-positive. Data are n/N (%).

Patient ID	Previous TB	Ultra semi-quantitation category	Programmatic Xpert or Ultra result*	Positive by eMRS and/or CRS	Treatment initiated after 12-week follow up
PCTB008	No	low	Not Done	Both	Yes
PCTB016	No	very low	Not Done	Both	Yes
PCTB023	No	trace	Not Done	Negative	No
PCTB047	No	very low	Xpert-negative	CRS	Yes
PCTB053	No	trace	Not Done	Both	Yes
PCTB058	No	trace	Not Done	Negative	No
PCTB068	Yes	trace	Not Done	Negative	No
PCTB069	Yes	trace	Not Done	CRS	Yes
PCTB071	No	trace	Not Done	Negative	No
PCTB072	No	trace	Not Done	CRS	Yes
PCTB075	No	trace	Not Done	Negative	No
PCTB079	No	very low	Not Done	Both	Yes
PCTB092	Yes	trace	Not Done	CRS	Yes
PCTB096	No	trace	Ultra-negative	CRS	Yes
PCTB101	Unknown	trace	Not Done	Negative	No
PCTB113	No	trace	Ultra-negative	Negative	No
PCTB117	No	trace	Ultra-negative	Negative	No
PCTB125	No	very low	Ultra-negative	Negative	No
PCTB126	No	trace	Not Done	Negative	No
PCTB127	No	trace	Ultra-negative	Negative	No
PCTB130	Yes	trace	Ultra-negative	CRS	Yes
PCTB134	No	very low	Ultra-positive (Low)	Negative	Unknown
PCTB136	No	trace	Ultra-negative	Negative	Unknown
PCTB137	No	very low	Ultra-positive (Trace)	Both	No
PCTB138	No	trace	Ultra-positive (Trace)	CRS	Yes
PCTB141	No	trace	Ultra-negative	Negative	Yes
PCTB142	No	very low	Ultra-negative	Negative	No
PCTB144	No	trace	Ultra-negative	CRS	Yes
PCTB146	No	trace	Not Done	Negative	No
PCTB154	No	very low	Ultra-positive (Low)	CRS	Yes
Overall	4/29 (14)	trace: 21/30 (70) very low: 8/30 (27) low: 1/30 (3)	Negative: 11/15 (73) Positive: 4/15 (27)	Both: 5/30 (17) eMRS alone: 0/30 (0) CRS alone: 9/30 (30) Negative: 16/30 (53)	14/28 (50)

*Not done because of the programmatic algorithm in **Figure 1**.

Abbreviations: CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Missing data: Previous TB, 1; treatment initiated, 2.

Supplementary Table 12: Non-head-to-head diagnostic accuracy of urinary Ultra (unconc. or conc.) and TB-LAM stratified by HIV status. Urine has low utility for diagnosing TBP. Ultra was done on concentrated urine and, if non-actionable, on an unconc. specimen. Urinary Ultra and TB-LAM had similar accuracy in head-to-head analyses (**Table 3**). Data are %, 95% CI, and n/N.

	All patients				**HIV-negative**				**HIV-positive**			
	n=98				**n=42/98 (43)**				**n=56/98 (57)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Ultra	26 (12, 43) 9/35	94 (85, 98) 59/63	69 (39, 91) 9/13	69 (58, 79) 59/85	9 (0, 41) 1/11	100 (89, 100) 31/31	100 (3, 100) 1/1	76 (60, 88) 31/41	33 (16, 55) 8/24 p=0.128*	88 (71, 96) 28/32 **p=0.042***	67 (35, 90) 8/12 p=0.488*	64 (48, 78) 28/44 p=0.231*
	n=79				**n=34/79 (43)**				**n=45/79 (57)**			
TB-LAM	27 (12, 46) 8/30 p=0.931‡	90 (78, 97) 44/49 p=0.457‡	62 (32, 86) 8/13 p=0.680‡	67 (54, 78) 44/66 p=0.719‡	9 (0, 41) 1/11 p>0.999‡	87 (66, 97) 20/23 **p=0.039‡**	25 (1, 81) 1/4 p=0.171‡	67 (47, 83) 20/30 p=0.408‡	37 (16, 62) 7/19 p=0.811‡ p=0.098*	92 (75, 99) 24/26 p=0.550‡ p=0.537*	78 (40, 97) 7/9 p=0.577‡ p=0.071*	67 (49, 81) 24/36 p=0.777‡ p>0.999*

Within column p-values: ‡ Unconcentrated and concentrated Ultra vs. TB-LAM.

Within row p-values: *HIV-negative vs. HIV-positive.

Abbreviations: CI, confidence interval; NPV, negative predictive value; PF, pericardial fluid; PPV, positive predictive value; TB-LAM, Determine TB-LAM, Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF

Supplementary Table 13: Comparison of characteristics by whether study unconcentrated Ultra was TP or FP per the MRS. A lower proportion of FPs were HIV-positive, started ART, experienced a fever, night sweats, had a pleural effusion, were on TB treatment at recruitment and after their 12-week follow-up, had serous-sanguineous PF or had a pericardial tamponade. FPs hence had milder disease. Additionally, FPs had higher *rpoB* and IS*6110*/IS*1081* C_{Ts} and were more likely to be trace-positive. Alternate reference standards (eMRS and CRS) detected TB in FPs. Data are n (%) or median (IQR).

	TPs (n=48)	**FPs** (n=30)
Clinical characteristics		
HIV-positive	35/48 (73)	9/30 (30) **p<0.001**
CD4 count (cells/µl)	127 (36-319)	223 (63-458) p=0.305
ART	9/35 (26)	5/8 (63) **p=0.045**
Pericardial tamponade	17/42 (40)	4/28 (14) **p=0.019**
Previous TB	14/47 (30)	4/29 (14) p=0.111
Current smoker	7/47 (15)	7/30 (23) p=0.349
Symptoms:		
Cough	29/44 (66)	17/29 (59) p=0.528
Fever	22/44 (50)	6/29 (21) **p=0.012**
Night sweats	28/45 (62)	8/28 (29) **p=0.005**
Weight loss	33/45 (73)	18/30 (60) p=0.225
Chest X-ray results:		
Normal	2/48 (4)	2/30 (7) p=0.626
Cardiomegaly	30/48 (63)	17/30 (57) p=0.609
Pulmonary infiltrates	13/48 (27)	4/30 (13) p=0.152
Hilar lymphadenopathy	3/48 (6)	3/30 (10) p=0.545

Miliary pattern	4/48 (8)	1/30 (3) p=0.380
Pleural effusion	35/48 (73)	14/30 (47) **p=0.020**
TB treatment		
Treatment at time of recruitment	12/48 (25)	2/30 (6) **p=0.040**
Treatment after 12-week follow-up	46/47 (98)	14/28 (50) **p<0.001**
Fluid characteristics		
Total PF aspirated (mL)	645 (500-908)	700 (538-700) p=0.237
Blood-stained	19/48 (40)	15/30 (50) p=0.367
Chylous	0/48 (0)	1/30 (3) p=0.203
Purulent	0/48 (0)	1/30 (3) p=0.203
Serous	15/48 (31)	10/30 (33) p=0.848
Serous-sanguineous	14/48 (29)	3/30 (10) **p=0.046**
Fluid cytochemistry		
Lymphocyte/ neutrophil ratio	1.37 (0.71-4.59)	1.36 (0.28-7.58) p=0.822
Adenosine deaminase (ADA; U/L)	62.10 (50.50-82.80)	20.90 (4.70-47.90) **p<0.001**
Total protein	58.00 (51.00-63.00)	51.00 (40.00-63.00) p=0.140
Lactate dehydrogenase (LDH)	1073 (554-1948)	593 (191-1146) **p=0.009**
Albumin	20.00 (16.00-24.00)	21.00 (0.00-29.00) p=0.555
Study unconcentrated Ultra result information		
rpoB C_{Tmin}	27.40 (22.15-29.75)	29.50 (28.90-35.50) **p=0.008**

IS*6110*/IS*1081* C_T	20.85 (18.33-23.15)	25.50 (23.23-27.25) **p<0.001**
Trace semi-quantitation category	11/48 (23)	21/30 (70) **p<0.001**
SPC C_T	25.30 (24.20-26.10)	25.00 (24.48-26.30) p=0.888
Programmatic and study Xpert and programmatic Ultra information		
Positive Xpert (unconcentrated and concentrated)	37/47 (79)	5/30 (17) **p<0.001**
Positive Ultra (concentrated)	13/13 (100)	4/15 (27) **p<0.001**
Study tests		
uIFN-γ	1495 (691-2883)	4 (0-1720) **p<0.001**
Alternate reference standards		
eMRS	48/48 (100)	5/30 (17) **p<0.001**
CRS	48/48 (100)	14/30 (47) **p<0.001**

Missing data: CD4 count, 1; ART, 1; pericardial tamponade, 8; previous TB,1; current smoker, 1; cough, 5; fever, 5; night sweats, 5; weight loss, 3; TB treatment after 12-week follow-up, 3; total PF aspirated, 8; lymphocyte/neutrophil ratio, 18; lymphocyte/neutrophil ratio undefined (divided by zero), 5; ADA, 4; total protein, 4; LDH, 4; albumin, 4; non-actionable Xpert results, 1; Ultras not done for true-positives (Xpert programmatically done), 35; Ultras not done for false-positives (Xpert programmatically done), 15.

Abbreviations: eMRS, extended microbiological reference standard; CRS, composite reference standard; FP, false-positive; C_T, cycle threshold; TP, true-positive; Ultra, Xpert MTB/RIF Ultra.

Supplementary Table 14: Diagnostic yield (proportion of people with at least one positive confirmatory test result detected by a test) of programmatic and study microbiological tests. PF Study Ultra had the highest diagnostic yield followed by culture on PF and culture on PF and non-site-of-disease fluid had higher yields in PLHIV. Data are n/N (%).

Test	Diagnostic yield			
	In patients who did or did not have the test attempted	**In people without HIV in patients who did or did not have the test attempted**	**In PLHIV in patients who did or did not have the test attempted**	**In patients who had the test attempted**
PF				
Study Ultra (unconc. and conc.)	91/109 (83)	40/47 (85)	51/62 (82) p=0.692	91/155 (59)
*Xpert (unconc. and conc.)	45/109 (41)	15/47 (32)	30/62 (48) p=0.084	45/152 (30)
Culture	57/109 (52)	17/47 (36)	40/62 (65) **p=0.003**	57/151 (38)
Pericardial biopsy				
Culture	15/109 (13)	6/47 (13)	9/62 (15) p=0.793	15/24 (63)
Non-site-of-disease fluid				
Smear microscopy	5/109 (5)	1/47 (2)	4/62 (6) p=0.285	5/5 (100)
Culture	33/109 (30)	9/47 (19)	24/62 (39) **p=0.028**	33/89 (37)
Programmatic Ultra	13/109 (12)	4/47 (9)	9/62 (15) p=0.338	13/56 (23)
Programmatic Xpert	10/109 (9)	3/47 (6)	7/62 (11) p=0.379	10/39 (26)
Urine				
Ultra (unconc. and conc.)	13/109 (12)	1/47 (2)	12/62 (19) **p=0.006**	13/99 (13)
TB-LAM	13/109 (12)	4/47 (9)	9/62 (15) p=0.338	13/79 (16)

*Includes programmatic concentrated Xpert and study unconcentrated Xpert (latter done by the study if former not done programmatically- only PF was used).

Abbreviations: Conc., concentrated; PF, pericardial fluid; MGIT960 culture, Mycobacteria Growth Indicator Tube 960; TB-LAM, Determine TB-LAM; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Appendix III

Site-of-disease Xpert MTB/RIF Ultra and urine tests for the diagnosis of tuberculosis pleuritis in an HIV-endemic setting

Chapter 4

(Supplementary material)

Table of Contents

Methods

Definitions

Actionable results

Xpert and Ultra: For TB, positive or negative. For rifampicin, TB-positive and either resistant or susceptible.

Culture: Positive or negative for MTBC.

Additional information of the programmatic diagnostic algorithm

The programmatic diagnostic algorithm differed based on if patients met the "previous PF or biopsy" definition of PF or biopsies collected and tested before the patient by the National Health Laboratory Service (NHLS) within three months prior to recruitment. If this was met and the previous samplings were Xpert- or Ultra-positive, 5-7.5 ml of the presently collected PF or biopsy was NALC-NaOH decontaminated, neutralised, centrifuged, and resuspended in phosphate buffer with 500 µl inoculated into a culture (**Figure 1)**. If these previous samplings were Xpert- or Ultra-negative, Xpert- or Ultra were repeated, together with culture, on the currently collected specimens. If patients did not meet this definition, diagnostic testing was done per the main text.

Results

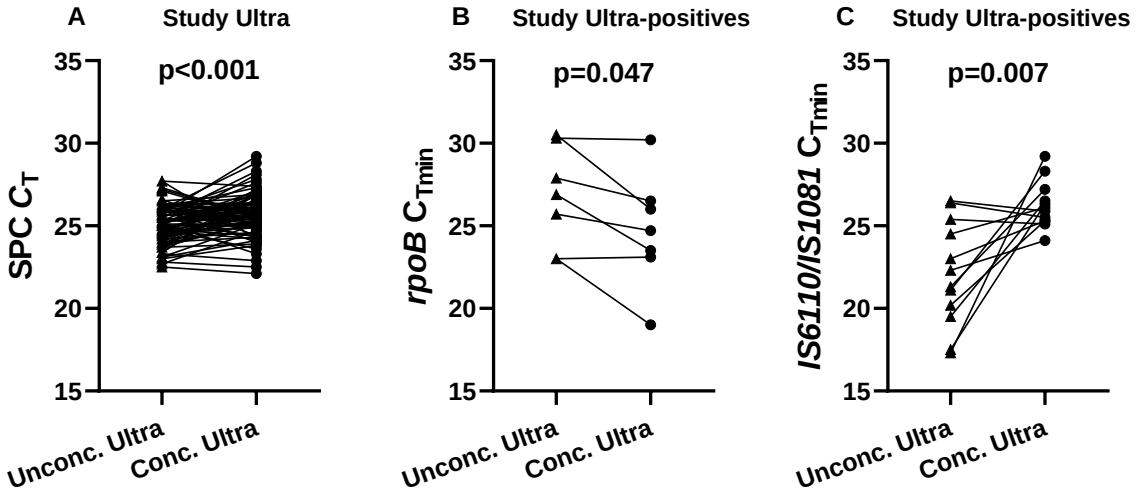
uIFN-γ on PF

Optimal cut-points: When specificity was prioritised (>5079.7 pg/ml), sensitivity 25% (95% confidence interval: 9, 48) and specificity was 95% (90, 98) and. At a cut-point of >1211.4 pg/ml (Youden's index), sensitivity was 94% (74, 100) and specificity 89% (81, 84).

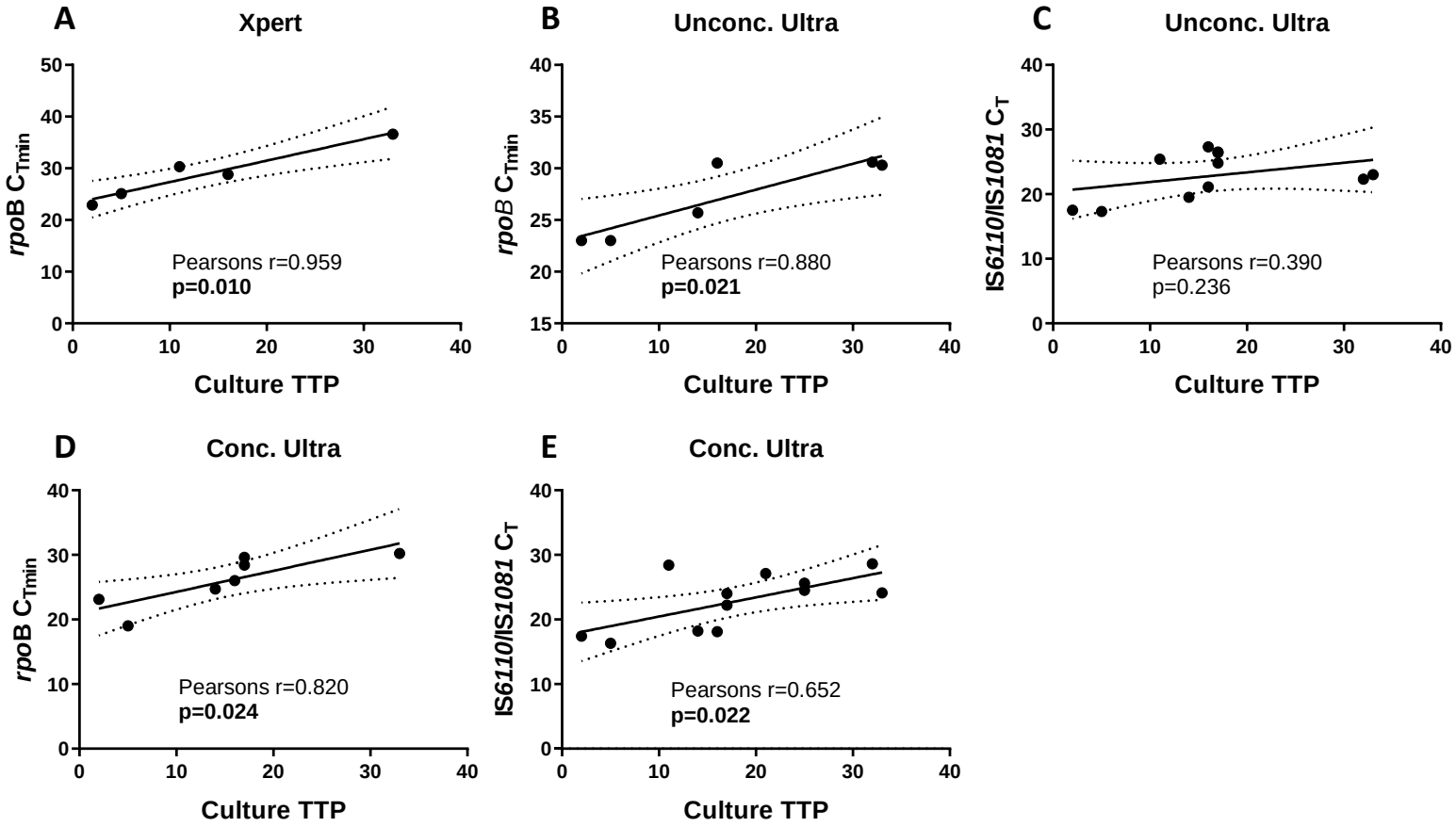
In PLHIV: The area under the curve (AUC) was 0.894 (95% confidence interval: 0.765, 1.000). Optimal rule-in, rule-out and Youden's index cut-points for uIFN-γ to maximise diagnostic accuracy were determined using the ROC curve shown in **Figure 4B**. At a rule-in cut-point (>5079.7 pg/ml), sensitivity was 33% (2, 87) and specificity was 95% (88, 98). At a rule-out cut-point (also Youden's index) (>1511.5 pg/ml), sensitivity was 100% (37, 100) and specificity was 90% (83, 95).

In people without HIV: The area under the curve (AUC) was 0.941 (0.879, 1.000). Optimal rule-in, rule-out and Youden's index cut-points for uIFN-γ to maximise diagnostic accuracy were determined using the ROC curve shown in **Figure 4C**. At a rule-in cut-point (>4354.5 pg/ml), sensitivity was 39% (17, 65) and specificity was 92% (68, 100). At a rule-out cut-point (>211.7 pg/ml), sensitivity was 100% (79, 100) and specificity 62% (36, 83). At a cut-point of >1211.4 pg/ml (Youden's index), sensitivity was 92% (68, 100) and specificity was 85% (59, 97).

Supplementary Figure 1: Spaghetti plots showing quantitative information (C_T) for different tests to measure the effect of concentration on PF. (**A**) PCR inhibitors (higher SPC C_T indicate more inhibition) and mycobacterial load measured using (**B**) the *rpoB* and **(C)** IS*6110*/IS*1081* probes (lower values indicate higher MTBC concentrations). Abbreviations: Conc., concentrated; PF, pleural fluid; C_T, cycle cut-point value; C_{Tmin}, minimum cycle cut-point value; SPC; sample processing control; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated.



Supplementary Figure 2: Quantitative information of Xpert and Ultra versus MGIT960 TTP. (**A**) shows Xpert whereas **(B-E)** shows unconc. and conc. Ultra results for each probe. A significant correlation between *rpo*B C_{Tmin} and TTP was observed regardless of concentration and, for Ultra IS*6110*/IS*1081* C_T, this was only observed after concentration. Xpert included programmatic and study results (and hence unconc. and conc. results) and, for Ultra, only study results are shown. Dashed lines indicate 95% confidence intervals. Abbreviations: Conc., concentrated; MGIT, mycobacteria growth indicator tube; PF, pleural fluid; TTP, culture time-to-positivity; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.



Supplementary Table 1: Reference standard definitions.

	MRS*	eMRS†	CRS‡
	Site of disease fluid		
MGIT960 on PF	✓	✓	✓
MGIT960 on pleural biopsy	✓	✓	✓
	Non-site-of disease fluid		
Smear microscopy	×	✓	✓
Xpert	×	✓	✓
Ultra	×	✓	✓
MGIT960	×	✓	✓
	Treatment information		
TB treatment initiated	×	×	✓
	Case definitions		
Reference standard **positive (Definite-TB cases)**	Any MRS test positive	Any eMRS test positive	Any eMRS test positive or TB treatment was initiated and no alternate diagnosis
Reference standard **negative (Non-TB patients)**	No MRS test positive	No eMRS test positive	No eMRS test positive and patient not initiated on treatment with and without an alternate diagnosis
Probable TB patients	N/A	N/A	N/A
Unclassifiable	No positive MRS test and site-of-disease fluid culture contaminated or not done	No positive eMRS test and site-of-disease fluid culture contaminated or not done	No positive eMRS test and site-of-disease fluid culture contaminated or not done, or treatment not initiated, or no alternate diagnosis available

Abbreviations: CRS, composite reference standard; eMRS, extended reference standard; MGIT960 culture, Mycobacteria Growth Indicator Tube 960 culture; MRS, microbiological reference standard; PF, pleural fluid; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 2: Sensitivity and specificity for different cut-off points of uIFN-γ concentrations stratified by HIV status. uIFN-γ had high sensitivity and specificity as a rule-out test and were similar when stratified by HIV status. Data are % and 95% CI.

	uIFN-γ								
	Overall			**HIV-positive**			**HIV-negative**		
	AUROC: 0.925 (0.873, 0.976)			**AUROC: 0.894 (0.765, 1.000)**			**AUROC: 0.941 (0.879, 1.000)**		
	Rule-in (>5079.7 pg/mL)	**Rule-out (>221.7 pg/mL)**	**Youden's index (>1211.4 pg/mL)**	**Rule-in (>5079.7 pg/mL)**	**Rule-out (>1511.5 pg/mL)**	**Youden's index (>1511.5 pg/mL)**	**Rule-in (>4354.5 pg/mL)**	**Rule-out (>211.7 pg/mL)**	**Youden's index (>1211.4 pg/mL)**
Sensitivity	25% (9, 48)	100% (83, 100)	94% (74, 100)	33% (2, 87)	100% (37, 100)	100% (37, 100)	39% (17, 65)	100% (79, 100)	92% (68, 100)
Specificity	95% (90, 98)	77% (68, 84)	89% (81, 94)	95% (88, 98)	90% (83, 95)	90% (83, 95)	92% (68, 100)	62% (36, 83)	85% (59, 97)

Abbreviations: AUROC, area under the receiver operator characteristic; uIFN-γ, unstimulated interferon-γ.

Supplementary Table 3: All non-actionable Xpert (study unconcentrated; information from programmatic testing unavailable) and study Ultra results (unconcentrated, concentrated). If a specimen had sufficient volume, the test was repeated once. The most common Xpert error was 5011: Signal loss detected in the amplification curve for analyte SPC and the most common error for concentrated Ultra was 2008: Syringe pressure reading the protocol limit. Data are % and n/N.

Test	Non-actionable rate	Patient ID	First result	Error code	Second result	Error code/reason test was not repeated	MRS status
Xpert (study unconc only)	5% (4/87)	PLTB045	Error	5011	Error	5007	Negative
		PLTB061	Error	5011	Not done	No more fluid	Negative
		PLTB065	Error	5007	No result	-	Unclassifiable
		PLTB088	Error	5011	Not done	No more fluid	Unclassifiable
0/2 (0%) non-actionable results resolved upon repeat testing							
Unconc. Ultra (study)	1% (1/114) p=0.093‡	PLTB079	Invalid	-	Not done	No more fluid	Unclassifiable
There was insufficient fluid for repeat testing							
Conc. Ultra (study)	12% (12/97) **p=0.001**± p=0.062¥	PLTB017	Invalid	-	Not done	No more fluid	Negative
		PLTB018	Invalid	-	Not done	No more fluid	Unclassifiable
		PLTB023	Error	2008	Not done	No more fluid	Negative
		PLTB033	Error	2008	Not done	No more fluid	Negative
		PLTB045	Error	2008	Not done	No more fluid	Negative
		PLTB047	Error	2008	Not done	No more fluid	Negative
		PLTB048	Error	2008	Not done	No more fluid	Negative
		PLTB049	Error	2008	Not done	No more fluid	Negative
		PLTB077	Error	2008	Not done	No more fluid	Negative
		PLTB079	Error	2008	Not done	No more fluid	Unclassifiable
		PLTB081	Error	2008	Negative	2008	Negative
		PLTB115	Error	2008	Not done	No more fluid	Negative
1/1 (100%) non-actionable results resolved upon repeat testing (1 Ultra-negative)							

Within column p-values: ‡ Study unconcentrated Xpert vs. study unconcentrated Ultra, ± study unconcentrated Ultra vs. study concentrated Ultra, ¥ study unconcentrated Xpert vs. study concentrated Ultra.

Abbreviations: Conc., concentrated; unconc., unconcentrated; MRS, microbiological reference standard

Error code definitions:
2008: Syringe pressure reading the protocol limit.
5011: Signal loss detected in the amplification curve for analyte SPC.
5007: SPC probe check failed.

Supplementary Table 4: Non-head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on PLF stratified by HIV status. Study unconcentrated Ultra had increased sensitivity and lower specificity overall compared to Xpert (specificity improved with concentration but sensitivity does not). uIFN-γ had similar sensitivity and specificity compared to study unconcentrated Ultra. The relative performances of all tests had similar patterns by HIV status. Conclusions were different to head-to-head analyses (**Table 2**). Data are %, 95% CI, and n/N.

	All patients				**HIV-negative**				**HIV-positive**			
	n=91				**n=68/91 (75)**				**n=23/91 (25)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	38 (14, 68) 5/13	97 (91, 100) 76/78	71 (29, 96) 5/7	90 (82, 96) 76/84	0 (0, 71) 0/3	98 (92, 100) 64/65	0 (0, 97) 0/1	96 (87, 99) 64/67	50 (19, 81) 5/10 p=0.119*	92 (64, 100) 12/13 p=0.200*	83 (36, 100) 5/6 p=0.088*	71 (44, 90) 12/17 **p=0.002***
	n=104				**n=77/103 (75)**				**n=26/103 (25)**			
Unconc. Ultra (study)	79 (49, 95) 11/14 **p=0.034‡**	70 (59, 79) 63/90 **p<0.001‡**	29 (15, 46) 11/38 **p=0.031‡**	95 (87, 99) 63/66 p=0.246‡	67 (9, 99) 2/3 p=0.083‡	68 (56, 78) 50/74 **p<0.001‡**	8 (1, 25) 2/26 p=0.773‡	98 (90, 100) 50/51 p=0.454‡	82 (48, 98) 9/11 p=0.122‡ p=0.571*	80 (52, 96) 12/15 p=0.353‡ p=0.340*	75 (43, 95) 9/12 p=0.689‡ **p<0.001***	86 (57, 98) 12/14 p=0.316‡ **p=0.052***
	n=80				**n=57/79 (72)**				**n=22/79 (28)**			
Conc. Ultra (study)	100 (74, 100) 12/12 p=0.088±	87 (76, 94) 59/68 **p=0.013±**	57 (34, 78) 12/21 **p=0.034±**	100 (94, 100) 59/59 p=0.097±	100 (16, 100) 2/2 p=0.361±	87 (76, 95) 48/55 **p=0.010±**	22 (3, 60) 2/9 p=0.238±	100 (93, 100) 48/48 p=0.330±	100 (69, 100) 10/10 p=0.156± p>0.999*	83 (52, 98) 10/12 p=0.825± p=0.717*	83 (52, 98) 10/12 p=0.615± **p=0.005***	100 (69, 100) 10/10 p=0.212± p>0.999*
	n=114				**n=82/113 (73)**				**n=31/113 (27)**			
uIFN-γ (rule-out cut-off 221.7 pg/ml)	83 (59, 96) 15/18 p=0.732¥	79 (70, 87) 76/96 p=0.151¥	43 (26, 61) 15/35 p=0.215¥	96 (89, 99) 76/79 p=0.822¥	75 (19, 99) 3/4 p=0.809¥	81 (70, 89) 63/78 p=0.063¥	17 (4, 41) 3/18 p=0.356¥	98 (92, 100) 63/64 p=0.871¥	86 (57, 98) 12/14 p=0.792¥ p=0.612*	71 (44, 90) 12/17 p=0.540¥ p=0.351*	71 (44, 90) 12/17 p=0.794¥ **p=0.001***	86 (57, 98) 12/14 p>0.999¥ **p=0.025***

Missing data: HIV,1.

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ±unconc. Ultra (study) vs. conc. Ultra (study), ¥unconc. Ultra (study) vs. uIFN-γ in patients of the same HIV status (overall, negative, positive).
Within row p-values: *HIV-negative vs. HIV-positive.

Abbreviations: Conc., concentrated; CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PLF, pleural fluid; PPV, positive predictive value; uIFN-γ, unstimulated interferon-γ; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Supplementary Table 5: Head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on PF compared to a MRS, eMRS and CRS for the detection of Mycobacterium tuberculosis complex DNA. Compared to the MRS, the CRS had decreased sensitivity for Xpert and Ultra (conc.) and increased specificity for uIFN-γ. The relative performance of eMRS compared to the MRS was the same. **(Table 2)**. Data are %, 95% CI, and n/N.

	MRS				**eMRS**				**CRS**			
	n=65				**n=65**				**n=65**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic unconc. and conc. pooled)	50 (19, 81) 5/10	98 (90, 100) 54/55	83 (36, 100) 5/6	92 (81, 97) 54/59	45 (17, 77) 5/11 p=0.835*	98 (90, 100) 53/54 p=0.990*	83 (36, 100) 5/6 p>0.999*	90 (79, 96) 53/59 p=0.752*	18 (6, 37) 5/28 p=0.076§ **p=0.048†**	97 (86, 100) 36/37 p=0.786§ p=0.775†	83 (36, 100) 5/6 p>0.999§ p>0.999†	61 (47, 73) 36/59 **p<0.001§** **p<0.001†**
Unconc. Ultra (study)	80 (44, 97) 8/10 p=0.160‡	65 (51, 78) 36/55 **p<0.001‡**	30 (14, 50) 8/27 **p=0.015‡**	95 (82, 99) 36/38 p=0.551‡	73 (39, 94) 8/11 p=0.193‡ p=0.696*	65 (51, 77) 35/54 **p<0.001‡** p=0.944*	30 (14, 50) 8/27 **p=0.015‡** p>0.999*	92 (79, 98) 35/38 p=0.706‡ p=0.644*	50 (31, 69) 14/28 **p=0.011‡** p=0.198§ p=0.099†	65 (47, 80) 24/37 **p<0.001‡** p=0.996§ p=0.954†	52 (32, 71) 14/27 p=0.158‡ p=0.097§ p=0.097†	63 (46, 78) 24/38 p=0.832‡ **p=0.003§** **p=0.001†**
Conc. Ultra (study)	100 (69, 100) 10/10 p=0.136±	85 (73, 94) 47/55 **p=0.015±**	56 (31, 78) 10/18 p=0.082±	100 (92, 100) 47/47 p=0.112±	91 (59, 100) 10/11 p=0.269± p=0.329*	85 (73, 93) 46/54 **p=0.015±** p=0.968*	56 (31, 78) 10/18 **p=0.082±** p>0.999*	98 (89, 100) 46/47 p=0.212± p=0.315*	54 (34, 72) 15/28 p=0.789± **p=0.029§** **p=0.008†**	92 (78, 98) 34/37 **p=0.005±** p=0.335§ p=0.351†	83 (59, 96) 15/18 **p=0.031±** p=0.070§ p=0.070†	72 (57, 84) 34/47 p=0.366± **p=0.001§** **p<0.001†**
uIFN-γ (rule-out cut-off 221.7 pg/ml))	100 (69, 100) 10/10 p=0.136¥	76 (63, 87) 42/55 p=0.208¥	43 (23, 66) 10/23 p=0.309¥	100 (92, 100) 42/42 p=0.132¥	91 (59, 100) 10/11 **p=0.269¥** p=0.329*	76 (62, 87) 41/54 p=0.206¥ p=0.957*	43 (23, 66) 10/23 p=0.309¥ p>0.999*	98 (87, 100) 41/42 p=0.258¥ p=0.314*	79 (59, 92) 22/28 **p=0.026¥** p=0.366§ p=0.111†	97 (86, 100) 36/37 p<0.001¥ **p=0.006§** **p=0.006†**	96 (78, 100) 22/23 **p=0.001¥** **p<0.001§** **p<0.001†**	86 (71, 95) 36/42 **p=0.020¥** **p=0.048§** **p=0.011†**

Missing data: Xpert non-actionable, n=2; Xpert not done, n=1; study concentrated Ultra non-actionable, n=7; study concentrated Ultra not done, n=3; uIFN-γ not done, n=3.
Within column p-values: ‡Xpert vs. unconc. Ultra (study), ±unconc. Ultra (study) vs. conc. Ultra (study), ¥unconc. Ultra (study) vs. uIFN-γ in patients using the same reference standard (MRS, eMRS, CRS).
Within row p-values: *MRS vs. eMRS, §eMRS vs. CRS, †MRS vs. CRS.

Abbreviations: conc., concentrated; CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PF, pleural fluid; PPV, positive predictive value; uIFN-γ, unstimulated interferon gamma; Ultra; Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Supplementary Table 6: The change in Ultra (unconc. and conc.) diagnostic accuracy on PF when Ultra traces were either excluded or reclassified as negative when compared to the MRS. Unconc. Ultra had similar sensitivity and decreased specificity when traces were excluded or reclassified. Conc. Ultra showed no change in diagnostic accuracy when traces were excluded but when traces were reclassified, sensitivity decreased and specificity increased. Data are %, 95% CI, and n/N.

	All patients				**HIV-negative**				**HIV-positive**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
	n=135				**n=62/135 (46)**				**n=73/135 (54)**			
Δ unconc. Ultra (study) traces excluded	+20% (-5, 45) p=0.334	+25% (10, 41) **p=0.006**	+28% (-13, 68) p=0.175	+5% (-18, 12) p=0.194	0%	+31% (13, 48) **p=0.005**	-5% (-15, 5) p=0.740	+3% (-3, 10) p=0.367	+12% (-10, 35) p=0.460	+0% (-29, 29) p>0.999	-8% (-49, 34) p=0.715	+11% (-9, 32) p=0.331
Δ conc. Ultra (study) traces excluded	0%	+9% (-3, 21) p=0.203	+11% (-33, 55) p=0.633	0%	0%	+11% (-2, 24) p=0.148	-22% (5, 49) p=0.598	0%	0%	0% (-29, 29) p>0.999	-9% (-50, 32) p=0.649	0%
Δ unconc. Ultra (study) traces reclassified	-30% (-70, 10) p=0.160	+29% (15, 42) **p<0.001**	+33% (-5, 71) p=0.091	-3% (-13, 7) p=0.515	-50% (-119, 19) p=0.248	+34% (19, 49) **p<0.001**	-5% (-15, 5) p=0.740	-1% (-10, 8) p=0.838	-25% (-66, 16) p=0.248	0% (-29, 29) p>0.999	-4% (-42, 33) p=0.826	-16% (-50, 17) p=0.369
Δ conc. Ultra (study) traces reclassified	-40% (-70, -10) **p=0.025**	+11% (0, 21) **p=0.047**	+19% (-18, 57) p=0.347	-7% (-13, 0) p=0.064	-50% (-119, 19) p=0.248	+13% (2, 24) **p=0.027**	+28% (-47, 102) p=0.425	-2% (-6, 2) p=0.354	-38% (-71, -4) p=0.055	0% (-29, 29) p>0.999	-6% (-42, 31) p=0.757	-27% (-54, -1) p=0.108

Abbreviations: Conc., concentrated; CI, confidence interval; MRS, microbiological reference standard; NPV, negative predictive value; PF, pleural fluid; PPV, positive predictive value; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated.

Supplementary Table 7: Rifampicin (RIF) results for Xpert (programmatic and study) and study Ultra (unconc. and conc.) on PF culture isolates that had programmatic RIF-susceptibility testing (MTBDR*plus*) done. Both Xpert and Ultra did not incorrectly classify any RIF-susceptible cases as RIF-resistant. Data %, 95% CI and n/N.

MTBDR*plus* result	**Xpert-positive***	**Unconc. Ultra-positive**	**Conc. Ultra-positive**
RIF-susceptible (11/11)	**RIF-resistant**		
	0% (0/3)	0% (0/7)	0% (0/7)
	RIF-susceptible		
	66% (2/3)	57% (4/7)	57% (4/7)
	RIF-indeterminate		
	33% (1/3)	43% (3/7)	43% (3/7)

*Xpert included study unconcentrated and programmatic concentrated results.

Missing data: Xperts not done=2, study unconc. Ultras not done=3, study conc. Ultras not done=4.

Abbreviations: conc., concentrated; MTBDR*plus*, GenoType MTBDR*plus*; PF, pleural fluid; RIF, rifampicin; unconc., unconcentrated.

Supplementary Table 8: Study Conc. and unconc. Ultra vs. programmatic conc. Ultra and study conc. Ultra vs. study unconc. Ultra on PF. Programmatic concentrated Ultras had more positive results compared to study unconcentrated Ultras.

A		**Conc. Ultra (study)**		
		Positive	Negative	Total
Conc. Ultra (programmatic)	Positive	5	3	8
	Negative	4	26	30
	Total	9	29	38
	Δ Conc. Ultra (study) vs. conc. Ultra (programmatic)	-3% (95% CI: -19, 14) p=0.706		
B		**Unconc. Ultra (study)**		
		Positive	Negative	Positive
Conc. Ultra (programmatic)	Positive	9	4	13
	Negative	12	28	40
	Total	21	32	53
	Δ Unonc. Ultra (study) vs. conc. Ultra (programmatic)	-15% (95% confidence interval; CI: -31, 1) **p=0.046**		
C		**Conc. Ultra (study)**		
		Positive	Negative	Positive
Unconc. Ultra (study)	Positive	16	19	35
	Negative	10	47	57
	Total	26	66	92
	Δ Unonc. Ultra (study) vs. conc. Ultra (study)	+10% (95% confidence interval; CI: -3, 22) p=0.095		

Non-actionable rates of programmatic conc. Ultras are unknown. Five study conc. Ultras were non-actionable (2 programmatically-positive, 3 negative) and study unconc. Ultra had no non-actionable results (repeats were done with sufficient PF per patient).

Abbreviations: CI, confidence interval; conc., concentrated; PF, pleural fluid; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated.

Supplementary Table 9: Comparison of characteristics by whether study unconcentrated Ultra was TP or FP per the MRS. More TPs were HIV-positive, showed a miliary pattern on their chest X-rays and were Xpert-, eMRS- and CRS-positive. FPs were more likely to have a pleural effusion, had higher IS*6110*/IS*1081* C_T and had lower uIFN-γ concentrations.

	TPs (n=11)	**FPs** (n=27)
Clinical characteristics		
HIV-positive	9/11 (82)	3/27 (11) **p<0.001**
CD4 count (cells/µl)	95 (28-238)	139 (78-302) p=0.617
ART	1/8 (13)	2/3 (67) p=0.072
Previous TB	4/10 (40)	4/27 (15) p=0.098
Current smoker	3/11 (27)	4/27 (15) p=0.369
Symptoms:		
Cough	6/10 (60)	13/26 (50) p=0.590
Fever	4/10 (40)	3/26 (12) p=0.053
Night sweats	6/10 (60)	8/26 (31) p=0.107
Weight loss	9/10 (90)	16/26 (62) p=0.097
Chest X-ray results:		
Normal	0/10 (0)	0/24 (0) p>0.999
Cardiomegaly	0/10 (0)	1/24 (4) p=0.512
Pulmonary infiltrates	1/10 (10)	4/24 (17) p=0.617
Hilar lymphadenopathy	0/10 (10)	0/24 (0) p>0.999
Miliary pattern	2/10 (20)	0/24 (0) **p=0.024**
Pleural effusion	8/10 (80)	24/24 (100) **p=0.024**
TB treatment		
Treatment after 12-week follow-up	8/11 (73)	10/26 (39)

		p=0.057
Fluid cytochemistry		
ADA (U/L)	47.90 (0-64.20)	18.65 (9.48-41.50) p=0.195
Total protein (g/L)	50.00 (37.00-62.00)	46.50 (27.00-54.25) p=0.336
LDH (U/L)	701 (371-1219)	421 (154-2083) p=0.349
Lymphocyte/ neutrophil ratio	3.65 (0.29-8.00)	4.00 (0.10-14.67) p=0.812
Study unconcentrated Ultra result information		
rpoB C_{Tmin}	28.00 (23.00-30.53)	27.05 (24.25-28.10) p=0.556
IS*6110*/IS*1081* C_T	23.00 (19.50-26.40)	27.00 (24.00-27.90) **p=0.011**
Trace semi-quantitation category	5/11 (45)	21/27 (78) p=0.052
SPC C_T	25.30 (23.90-26.00)	25.20 (24.10-26.00) p=0.830
Programmatic and study Xpert and programmatic Ultra information		
Positive Xpert (unconc. and conc.)	5/10 (50)	2/23 (9) **p=0.008**
Positive Ultra (conc.)	2/5 (40)	4/13 (31) p=0.710
Study tests		
uIFN-γ	4432 (2359-6129)	0 (0-435) **p<0.001**
Alternate reference standards		
eMRS	11/11 (100)	0/27 (0) **p<0.001**
CRS	11/11 (100)	10/27 (37) **p<0.001**

Missing data: CD4 count, 2; ART, 1; previous TB,1; cough, 2; fever, 2; night sweats, 2; weight loss, 2; chest X-ray, 4; TB treatment (unsuccessful follow-up), 1; ADA, 1; total protein, 1; LDH, 1; lymphocyte/neutrophil ratio, 12; Xpert,4; programmatic Ultras, 20.

Abbreviations: ADA, Adenosine deaminase; conc., concentrated; CRS, composite reference standard; C_T, cycle cut-point; eMRS, extended microbiological reference standard; FP, false-positive; LDH, lactate dehydrogenase; TP, true-positive; uIFN-γ; unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated.

Supplementary Table 10: Per patient information for study unconcentrated Ultra-positive patients (on PF) that were MRS-negative, with previous TB status, Ultra semi-quantitation category, programmatic Xpert or Ultra results, alternate reference standard result and treatment initiation status after at least 12-weeks follow-up. Approximately 4 in 5 MRS-negative Ultra-positives were trace-positive. Data are n/N (%).

Patient ID	Previous TB	Ultra semi-quantitation category	Programmatic Xpert or Ultra result*	Positive by eMRS and/or CRS	Treatment initiated after 12-week follow up
PLTB024	No	Trace	Not done	CRS	Yes
PLTB028	No	Trace	Not done	Negative	No
PLTB032	No	Trace	Negative	Negative	No
PLTB035	No	Trace	Positive	Negative	No
PLTB038	No	Trace	Non-actionable	Negative	No
PLTB039	No	Trace	Not done	Negative	No
PLTB042	Yes	Trace	Not done	CRS	Yes
PLTB045	No	Low	Negative	CRS	Yes
PLTB047	No	Trace	Not Done	CRS	Yes
PLTB050	No	Trace	Not Done	Negative	No
PLTB063	No	Trace	Not done	CRS	Yes
PLTB064	No	Trace	Not done	CRS	Yes
PLTB066	No	Low	Not Done	Negative	No
PLTB077	No	Trace	Positive	CRS	Yes
PLTB081	No	Trace	Negative	Negative	No
PLTB083	No	Low	Negative	Negative	No
PLTB084	No	Low	Negative	Negative	No
PLTB086	No	Trace	Not Done	Negative	No
PLTB091	No	Trace	Negative	Negative	No
PLTB092	No	Low	Not Done	Negative	No
PLTB094	Yes	Trace	Negative	Negative	No
PLTB097	No	Trace	Positive	CRS	Yes
PLTB102	No	Trace	Negative	CRS	Yes
PLTB103	Yes	Trace	Not done	Negative	Unsuccessful follow up after 3 calls
PLTB107	No	Trace	Not done	CRS	Yes
PLTB115	Yes	Trace	Positive	Negative	No
PLTB117	No	Trace	Negative	Negative	No
Overall	4/27 (15)	trace: 22/27 (81) low: 5/27 (19)	Negative: 9/14 (64) Positive: 4/14 (29) Non-actionable: 1/14 (7)	Both: 0/27 (0) eMRS alone: 0/27 (0) CRS alone: 10/27 (37) Negative: 17/27 (63)	Yes:10/26 (38) No: 16/26 (62)

*Not done because of the programmatic algorithm in **Figure 1**.

Abbreviations: CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; PF, pleural fluid; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Missing data: Treatment initiated, 1.

Supplementary Table 11 Diagnostic yield (proportion of people with at least one positive confirmatory test result detected by a test) of programmatic and study microbiological tests. PF Study Ultra had the highest diagnostic yield followed by culture on PF. Xpert and culture on PF and smear microscopy, culture and programmatic Ultra on non-site-of-disease fluid had higher yields in people without HIV. Data are n/N (%).

Test	Diagnostic yield			
	In patients who did or did not have the test attempted	**In people without HIV in patients who did or did not have the test attempted**	**In PLHIV in patients who did or did not have the test attempted**	**In patients who had the test attempted**
PF				
Study Ultra (unconc. and conc.)	51/65 (78)	15/20 (75)	36/45 (80) p=0.651	51/113 (45)
*Xpert (unconc. and conc.)	8/65 (12)	6/20 (30)	2/45 (4) **p=0.004**	8/98 (8)
Culture	18/65 (28)	14/20 (70)	4/45 (9) **p<0.001**	18/121 (15)
Pleural biopsy				
Culture	0/65 (0)	0/20 (0)	0/45 (0) p>0.999	0/3 (0)
Non-site-of-disease fluid				
Smear microscopy	3/65 (5)	3/20 (15)	0/45 (0) **p=0.008**	3/3 (100)
Culture	6/65 (9)	4/20 (20)	2/45 (4) **p=0.046**	6/31 (19)
Programmatic Ultra	8/65 (12)	5/20 (25)	3/45 (7) **p=0.038**	8/32 (25)
Programmatic Xpert	1/65 (2)	1/20 (9)	0/45 (0) p=0.131	1/3 (33)
Urine				
Ultra (unconc. and conc.)	4/65 (6)	2/20 (10)	2/45 (4) p=0.390	4/35 (11)
TB-LAM	8/65 (12)	3/20 (15)	5/45 (11) p=0.660	8/32 (25)

*Includes programmatic concentrated Xpert and study unconcentrated Xpert (latter done by the study if former not done programmatically- only PF was used).

Abbreviations: Conc., concentrated; PF, pleural fluid; MGIT960 culture, Mycobacteria Growth Indicator Tube 960; TB-LAM, Determine TB-LAM; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Appendix IV

Xpert MTB/RIF Ultra accurately diagnoses pulmonary tuberculosis using bronchial fluid in an HIV-endemic setting

Chapter 5

(Supplementary material)

Table of Contents

Supplementary Method

Additional information of the programmatic diagnostic algorithm

Current BALF, BWF or biopsy: BALF or BWF or biopsies collected as part of the study (after patient recruitment) are referred to as current BALF or BWF (**Figure 1**).

Previous BALF, BWF or biopsy: Different BALF or BWF or transbronchial lung biopsies may have been collected and tested within 3 months of recruitment through the government laboratory [National Health Laboratory Service (NHLS)], referred to as previous BALF or BWF or biopsies.

Use of previous BALF, BWF or biopsy: The previous BALF, BWF or biopsy Ultra results were used for NHLS downstream testing purposes. If a previous BALF, BWF or biopsy was Xpert- or Ultra-positive, then current BALF, BWF or biopsy was decontaminated for MGIT960 culture only. If a previous BALF, BWF or biopsy was Xpert or Ultra-negative, then current BALF, BWF or biopsy was decontaminated for MGIT culture and Xpert or Ultra testing. If a previous BALF, BWF or biopsy was not collected, the current BALF, BWF or biopsy was used for all tests.

Supplementary Results

uIFN-γ on BALF/BWF

Optimal cut-points

In PLHIV: The area under the curve (AUC) was 0.627 (95% confidence interval: 0.366, 0.887). Optimal cut-points for uIFN-γ on BALF and BWF were determined using a ROC curve (**Figure 4B)**. When we prioritised sensitivity (≥0.0 pg/ml), sensitivity was 100% (55, 100) and specificity was 0% (0, 6). When specificity was prioritised (≥7.8 pg/ml), sensitivity was 20% (1, 66) and specificity was 96% (87, 99). For youden's index (≥0.7 pg/ml), the sensitivity and specificity was 60% (19, 92) and 69% (56, 80) respectively.

In people without HIV: The area under the curve (AUC) was 0.623 (0.510, 0.735). Optimal cut-points for uIFN-γ on BALF and BWF were determined using a ROC curve (**Figure 4C)**. When we prioritised sensitivity (≥0.0 pg/ml), sensitivity was 100% (91, 100) and specificity was 0% (0, 1). When specificity was prioritised (≥4.1 pg/ml), sensitivity was 24% (13, 40) and specificity was 95% (92, 97). For youden's index (≥0.1 pg/ml), the sensitivity and specificity was 46% (31, 61) and 76% (71, 80) respectively.

BALF/BWF results

Ultra PCR inhibition: An analysis of sample processing control (SPC) C_T values showed more inhibition in study concentrated Ultras than study unconcentrated Ultras [25.90 (IQR: 24.90-26.90) vs. 25.10 (24.40-26.10); p<0.001], suggesting inhibition in concentrated Ultras contributes to diminished sensitivity. (**Supplementary Figure 1 A**; higher C_T values indicate more inhibition).

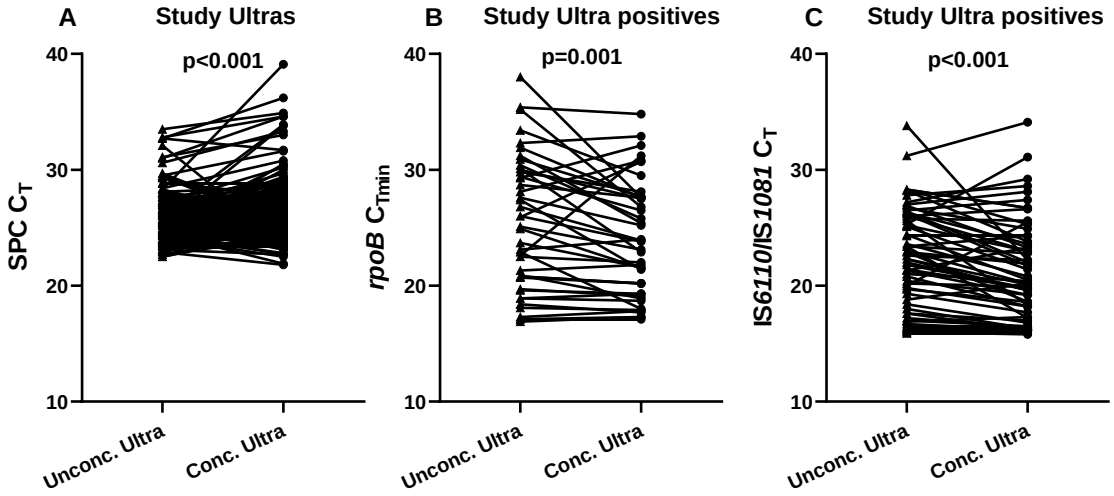
Ultra PCR rpoB $C_T$min and IS6110/IS1081 C_T: An analysis of *rpo*B $C_T$min and IS*6110*/IS*1081* C_T values showed that study concentrated Ultra had lower *rpo*B $C_T$min compared to unconcentrated Ultra [23.10 (IQR: 19.30-27.60) vs. 25.10 (20.70-30.00);

p=0.001] and lower IS*6110*/IS*1081* C_T [19.90 (16.50-23.45) vs. 22.00 (17.40-25.90); p<0.001] values (**Supplementary Figure 1B and 1C**; lower C_T values indicates a higher bacterial load); suggesting that study concentrated Ultras had a higher *Mycobacterium tuberculosis* (*M.tb)* bacillary load.

Relationship with bacterial load: *rpo*B $C_T$min showed a correlation with bacillary load (measured using culture time-to-positivity; TTP) in Xpert and both study unconcentrated Ultra and study concentrated Ultra on BALF/BWF (**Supplementary Figure 2A, 2B and 2D**). IS*6110*/IS*1081* C_T showed a correlation with bacillary load in study unconcentrated Ultras on BALF/BWF (**Supplementary Figure 2C**).

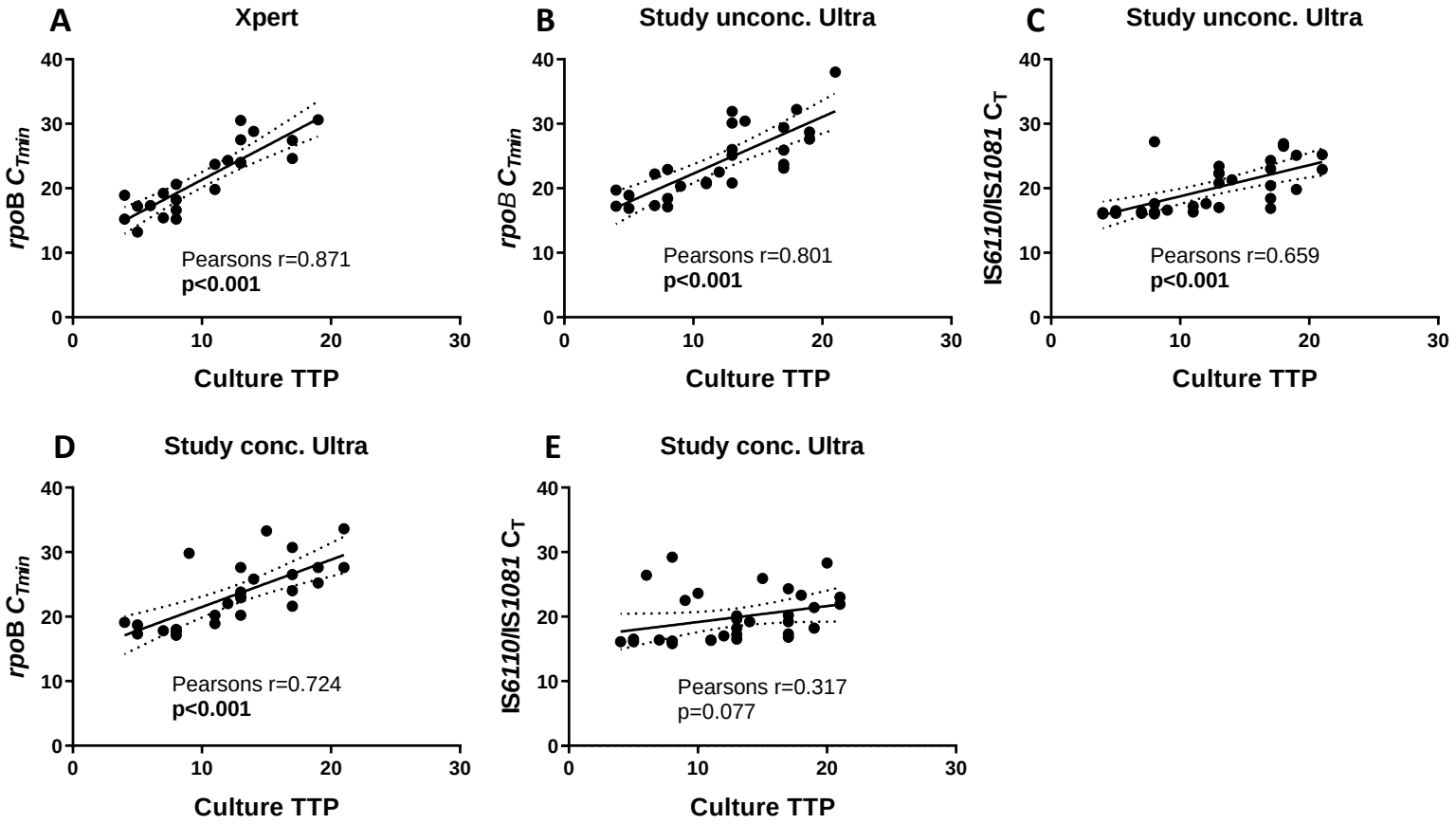
Drug-susceptibility testing: Of the isolates that had programmatic MTBDR*plus* (molecular line probe assay) testing done, Xpert (programmatic concentrated and study unconcentrated) and study unconcentrated Ultra had no false-negative or false-positive rifampicin results. Study concentrated Ultra had one false-positive rifampicin result (**Supplementary Table 9**).

Supplementary Figure 1: Spaghetti plots showing PCR quantitative information (C_T) for different tests to measure the effect of concentration on BALF or BWF. (**A**) PCR inhibitors (higher SPC C_T indicate more inhibition) and mycobacterial load measured using (**B**) the *rpoB* and **(C)** IS*6110*/IS*1081* probes (lower values indicate higher MTBC concentrations). Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; C_T, cycle threshold value; C_{Tmin}, minimum cycle threshold value; SPC; sample processing control; Ultra, Xpert MTB/RIF Ultra; unconc.,unconcentrated.



*Six study concentrated Ultras and two study unconcentrated Ultras SPC values were excluded (SPC=0).

Supplementary Figure 2: Quantitative information of Xpert and Ultra versus MGIT960 TTP. (**A**) Programmatic and study Xpert *rpo*B C_{Tmin} vs. MGIT960 liquid culture TTP. (**A**) shows Xpert whereas **(B-E)** shows unconc. and conc. Ultra results for each probe. A significant correlation was observed between *rpo*B C_{Tmin} and TTP regardless of concentration and IS*6110*/IS*1081* C_T and TTP in unconcentrated Ultras. Xpert included programmatic and study results (and hence unconc. and conc. results) and, for Ultra, only study results are shown. Dashed lines indicate 95% confidence intervals. Abbreviations: Conc., concentrated; MGIT, mycobacteria growth indicator tube; TTP, culture time-to-positivity; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.



Supplementary Table 1: Reference standard definitions

	MRS*	eMRS†	CRS‡
	Site of disease fluid		
MGIT960 culture on BALF or BWF	✓	✓	✓
MGIT960 culture on lung biopsy	✓	✓	✓
	Non-site-of disease fluid		
Smear microscopy	×	✓	✓
Xpert	×	✓	✓
Ultra	×	✓	✓
MGIT960 Culture	×	✓	✓
	Alternate diagnosis and treatment information		
Alternate diagnosis	×	×	✓
TB treatment initiated	×	×	✓
	Case definitions		
Reference standard **positive (Definite-TB cases)**	Any MRS test positive	Any eMRS test positive	Any eMRS test positive or TB treatment was initiated and no alternate diagnosis
Reference standard **negative (Non-TB patients)**	No MRS test positive	No eMRS test positive	No eMRS test positive and patient not initiated on treatment with and without an alternate diagnosis
Probable TB patients	N/A	N/A	N/A
Unclassifiable	No positive MRS test and site-of-disease fluid culture contaminated or not done	No positive eMRS test and site-of-disease fluid culture contaminated or not done	No positive eMRS test and site-of-disease fluid culture contaminated or not done, and treatment not initiated, and no alternate diagnosis available

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 2: Sensitivity and specificity for different cut-off points of uIFN-γ concentrations stratified by HIV status. Sensitivity and specificity for uIFN-γ cut-points were similar when stratified by HIV status. Data are % and 95% CI.

	uIFN-γ								
	Overall			**HIV-negative**			**HIV-positive**		
	AUROC: 0.624 (0.520, 0.723)			**AUROC: 0.623 (0.510, 0.735)**			**AUROC: 0.627 (0.366, 0.887)**		
	Rule-in (≥4.2 pg/mL)	**Rule-out (≥0.0 pg/mL)**	**Youden's index (≥0.1 pg/mL)**	**Rule-in (≥4.1 pg/mL)**	**Rule-out (≥0.0 pg/mL)**	**Youden's index (≥0.1 pg/mL)**	**Rule-in (≥7.8 pg/mL)**	**Rule-out ≥0.0 pg/mL)**	**Youden's index (≥0.7 pg/mL)**
Sensitivity	24 (13, 38)	100 (92, 100)	47 (33, 62)	24 (13, 40)	100 (91, 100)	46 (31, 61)	20 (1, 66)	100 (55, 100)	60 (19, 92)
Specificity	95 (92, 97)	0 (0, 1)	64 (55, 73)	95 (92, 97)	0 (0, 1)	76 (71, 80)	96 (87, 99)	0 (0, 6)	69 (56, 80)

Abbreviations: AUROC, area under the receiver operator characteristic; uIFN-γ, unstimulated interferon-γ.

Supplementary Table 3: Head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc.and conc.) and uIFN-γ on BALF and BWF versus the MRS stratified by fluid type. Diagnostic accuracy was similar between BALF and BWF specimen collections. Data are %, 95% CI, and n/N.

	All patients (n=276)				**Bronchoalveolar lavage fluid (n=124)**				**Bronchial wash fluid (n=152)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	59 (41, 75) 20/34	96 (93, 98) 232/242	67 (47, 83) 20/30	94 (91, 97) 232/246	46 (19, 75) 6/13	95 (89, 98) 105/111	50 (21, 79) 6/12	94 (88, 97) 105/112	67 (43, 85) 14/21 p=0.238	97 (92, 99) 127/131 p=0.360	78 (52, 94) 14/18 p=0.114	95 (90, 98) 127/134 p=0.730
Unconc. Ultra (study)	79 (62, 91) 27/34 p=0.066‡	72 (66, 77) 174/242 **p<0.001**‡	28 (20, 39) 27/95 **p<0.001**‡	96 (92, 98) 174/181 p=0.389‡	69 (39, 91) 9/13 p=0.234‡	68 (59, 77) 76/111 **p<0.001**‡	20 (10, 35) 9/44 **p=0.041**‡	95 (88, 99) 76/80 p=0.713‡	86 (64, 97) 18/21 p=0.147‡ p=0.248	75 (66, 82) 98/131 **p<0.001**‡ p=0.274	35 (22, 50) 18/51 **p=0.002**‡ p=0.110	97 (92, 99) 98/101 p=0.397‡ p=0.482
Conc. Ultra (study)	91 (76, 98) 31/34 p=0.171± **p=0.002**§	77 (71, 82) 187/242 p=0.179± **p<0.001**§	36 (26, 47) 31/86 p=0.272± **p=0.004**§	98 (95, 100) 187/190 p=0.174± **p=0.028**§	92 (64, 100) 12/13 p=0.136± p=0.234§	77 (69, 85) 86/111 p=0.131± **p<0.001**§	32 (18, 50) 12/37 p=0. 220± p=0273§	99 (94, 100) 86/87 p=0.145± p=0.069§	90 (70, 99) 19/21 p=0.634± p=0.060§ p=0.855	77 (69, 84) 101/131 p=0.665± **p<0.001**§ p=0.944	39 (25, 54) 19/49 p=0.719± **p=0.005**§ p=0.544	98 (93, 100) 101/103 p=0.635± p=0.190§ p=0.635
uIFN-γ (rule-in cut-off 4.2 pg/ml)	18 (7, 35) 6/34 **p<0.001**¥	96 (93, 98) 232/242 **p<0.001**¥	38 (15, 65) 6/16 p=0.462¥	89 (85, 93) 232/260 **p=0.039**¥	23 (5, 54) 3/13 **p=0.018**¥	93 (86, 97) 103/111 **p<0.001**¥	27 (6, 61) 3/11 p=0.624¥	91 (84, 96) 103/113 p=0.310¥	14 (3, 36) 3/21 **p<0.001**¥ p=0.514	98 (95, 100) 129/131 **p<0.001**¥ **p=0.027**	60 (15, 95) 3/5 p=0.276¥ p=0.210	88 (81, 93) 129/147 **p=0.010**¥ p=0.381

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ± unconc. Ultra (study) vs. conc. Ultra (study), §Xpert vs. conc. Ultra (study), Ψ unconc. Ultra (study) vs. uIFN-γ in patients with the same fluid type (overall, bronchoalveolar lavage, bronchial wash).

Within row p-values: Bronchoalveolar lavage fluid vs. bronchial wash fluid.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; CI, confidence interval; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; unconc., unconcentrated; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 4: Non-head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on BALF and BWF versus the MRS stratified by fluid type. Trends were like the head-to-head analysis (**Supplementary Table 2**). Data are %, 95% CI, and n/N.

	All patients (n=320)				**Bronchoalveolar lavage fluid (n=134)**				**Bronchial wash fluid (n=186)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	58 (41, 74) 22/38	95 (91, 97) 267/282	59 (42, 75) 22/37	94 (91, 97) 267/283	46 (19, 75) 6/13	94 (88, 98) 114/121	46 (19, 75) 6/13	94 (88, 98) 114/121	64 (43, 82) 16/25 p=0.291	95 (90, 98) 153/161 p=0.762	67 (45, 84) 16/24 p=0.225	94 (90, 97) 153/162 p=0.934
	n=324				**n=134**				**n=190**			
Unconc. Ultra (study)	76 (60, 88) 31/41 p=0.094‡	71 (65, 76) 200/283 **p<0.001**‡	27 (19, 36) 31/114 **p<0.001**‡	95 (91, 98) 200/210 p=0.661‡	69 (39, 91) 9/13 p=0.234‡	69 (60, 77) 83/121 **p<0.001**‡	19 (9, 33) 9/47 **p=0.047**‡	95 (89, 99) 83/87 p=0.706‡	79 (59, 92) 22/28 p=0.240‡ p=0.517	72 (65, 79) 117/162 **p<0.001**‡ p=0.507	33 (22, 45) 22/67 **p=0.002**‡ p=0.106	95 (90, 98) 117/123 p=0.800‡ p=0.925
	n=289				**n=129**				**n=160**			
Conc. Ultra (study)	89 (74, 97) 32/36 p=0.132± **p=0.003**§	77 (71, 82) 195/253 p=0.093± **p<0.001**§	36 (26, 46) 32/90 p=0.199± **p=0.0113**§	98 (95, 99) 195/199 **p=0.048**± p=0.126§	92 (64, 100) 12/13 p=0.136± **p=0.011**§	77 (68, 84) 89/116 p=0.161± **p<0.001**§	31 (17, 48) 12/39 p=0.212± p=0.131§	99 (94, 100) 89/90 p=0.162± p=0.079§	87 (66, 97) 20/23 p=0.434± p=0.067§ p=0.624	77 (69, 84) 106/137 p=0.308± **p<0.001**§ p=0.903	39 (26, 54) 20/51 p=0.473± **p=0.026**§ p=0.407	97 (92, 99) 106/109 p=0.403± p=0.271§ p=0.412
	n=323				**n=134**				**n=189**			
uIFN-γ (rule-in cut-off 4.2 pg/ml)	23 (11, 38) 9/40 **p<0.001**¥	95 (92, 97) 269/283 **p<0.001**¥	39 (20, 61) 9/23 p=0.251¥	90 (86, 93) 269/300 **p=0.023**¥	23 (5, 54) 3/13 **p=0.018**¥	93 (87, 97) 113/121 **p<0.001**¥	27 (6, 61) 3/11 p=0.549¥	92 (86, 96) 113/123 p=0.312¥	22 (9, 42) 6/27 **p<0.001**¥ p=0.952	96 (92, 99) 156/162 **p<0.001**¥ p=0.264	50 (21, 79) 6/12 p=0.252¥ p=0.265	88 (82, 93) 156/177 **p=0.038**¥ p=0.296

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ± unconc. Ultra (study) vs. conc. Ultra (study), §Xpert vs. conc. Ultra (study), ¥ unconc. Ultra (study) vs. uIFN-γ in patients with the same fluid type (overall, bronchoalveolar lavage, bronchial wash).

Within row p-values: Bronchoalveolar lavage fluid vs. bronchial wash fluid.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; CI, confidence interval; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; unconc., unconcentrated; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 5: Non-head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on BALF and BWF versus the MRS stratified by HIV status. Trends were like the head-to-head analysis (**Table 2**). Data are %, 95% CI, and n/N.

	All patients (n=320)				**HIV-negative (n=269)**				**HIV-positive (n=51)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	58 (41, 74) 22/38	95 (91, 97) 267/282	59 (42, 75) 22/37	94 (91, 97) 267/283	64 (45, 80) 21/33	94 (91, 97) 223/236	62 (44, 78) 21/34	95 (91, 97) 223/235	20 (1, 72) 1/5 p=0.066	96 (85, 99) 44/46 p=0.748	33 (1, 91) 1/3 p=0.336	92 (80, 98) 44/48 p=0.378
	n=324				**n=273**				**n=51**			
Unconc. Ultra (study)	76 (60, 88) 31/41 p=0.094‡	71 (65, 76) 200/283 **p<0.001‡**	27 (19, 36) 31/114 **p<0.001‡**	95 (91, 98) 200/210 p=0.661‡	81 (64, 92) 29/36 p=0.116‡	71 (65, 77) 169/237 **p<0.001‡**	30 (21, 40) 29/97 **p=0.001‡**	96 (92, 98) 169/176 p=0.590‡	40 (5, 85) 2/5 p=0.490‡ p=0.048	67 (52, 80) 31/46 **p=0.001‡** p=0.593	12 (1, 36) 2/17 p=0.335‡ p=0.121	91 (76, 98) 31/34 p=0.938‡ p=0.226
	n=289				**n=247**				**n=42**			
Conc. Ultra (study)	89 (74, 97) 32/36 p=0.132± **p=0.003§**	77 (71, 82) 195/253 p=0.093± **p<0.001§**	36 (26, 46) 32/90 p=0.199± **p=0.0113§**	98 (95, 99) 195/199 **p=0.048±** p=0.126§	91 (75, 98) 29/32 p=0.242± **p=0.010§**	78 (72, 83) 168/215 p=0.096± **p<0.001§**	38 (27, 50) 29/76 p=0.253± **p=0.022§**	98 (95, 100) 168/171 p=0.216± p=0.077§	75 (19, 99) 3/4 p=0.294± p=0.099§ p=0.349	71 (54, 85) 27/38 p=0.718± **p=0.002§** p=0.338	21 (5, 51) 3/14 p=0.467± p=0.659§ p=0.230	96 (82, 100) 27/28 p=0.402± p=0.419§ p=0.525
	n=323				**n=273**				**n=50**			
uIFN-γ (rule-in cut-off 4.2 pg/ml)	23 (11, 38) 9/40 **p<0.001¥**	95 (92, 97) 269/283 **p<0.001¥**	39 (20, 61) 9/23 p=0.251¥	90 (86, 93) 269/300 **p=0.023¥**	23 (10, 40) 8/35 **p<0.001¥**	96 (92, 98) 228/238 **p<0.001¥**	44 (22, 69) 8/18 p=0.225¥	89 (85, 93) 228/255 **p=0.012¥**	20 (1, 72) 1/5 p=0.490¥ p=0.886	91 (79, 98) 41/45 **p=0.005¥** p=0.184	20 (1, 72) 1/5 p=0.637¥ p=0.322	91 (79, 98) 41/45 p=0.992¥ p=0.730

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ± unconc. Ultra (study) vs. conc. Ultra (study), §Xpert vs. conc. Ultra (study), ¥ unconc. Ultra (study) vs. uIFN-γ in patients with the same fluid type (overall, bronchoalveolar lavage, bronchial wash).

Within row p-values: HIV-negative vs. HIV-positive.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; CI, confidence interval; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; unconc., unconcentrated; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 6: The change in Ultra (unconc. and conc.) diagnostic accuracy on BALF and BWF when Ultra traces were either excluded or reclassified as negative when compared to the MRS. Both unconc. and conc. Ultra had increased specificity when traces were excluded or reclassified as negative, while sensitivity remained similar. Trace exclusion and reclassification were similar in people living with HIV. Data are %, 95% CI, and n/N.

	All patients				**HIV-negative**				**HIV-positive**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
	n=135				**n=62/135 (46)**				**n=73/135 (54)**			
Δ unconc. Ultra (study) traces excluded	-3% (-22, 17) p=0.787	+19% (12, 26) **p<0.001**	+31% (13, 48) **p=0.001**	+1% (-3, 5) p=0.702	+1% (-16, 19) p=0.883	+19% (11, 27) **p<0.001**	+33% (14, 52) **p=0.001**	0% (-3, 4) p=0.822	+8% (-60, 76) p=0.809	+18% (-3, 38) p=0.108	+12% (-26, 50) p=0.457	+3% (-13, 18) p=0.736
Δ conc. Ultra (study) traces excluded	-2% (-17, 13) p=0.802	+12% (4, 19) **p=0.002**	+21% (3, 39) **p=0.024**	0% (-3, 2) p=0.807	-1% (-15, 13) p=0.850	+12% (4, 19) **p=0.003**	+23% (4, 43) **p=0.021**	0% (-3, 2) p=0.837	-8% (-76, 60) p=0.809	+11% (-10, 32) p=0.306	+7% (-33, 47) p=0.717	-1% (-12, 10) p=0.908
Δ unconc. Ultra (study) traces reclassified	-2% (-17, 13) p=0.802	+12% (4, 19) **p=0.002**	+21% (3, 39) **p=0.024**	0% (-3, 2) p=0.807	-1% (-15, 13) p=0.850	+12% (4, 19) **p=0.003**	+23% (4, 43) **p=0.021**	0% (-3, 2) p=0.837	-8% (-76, 60) p=0.809	+11% (-10, 32) p=0.306	+7% (-33, 47) p=0.717	-1% (-12, 10) p=0.908
Δ conc. Ultra (study) traces reclassified	-15% (-32, 2) p=0.100	+12% (5, 19) **p<0.001**	+14% (-3, 31) p=0.107	-2% (-5, 1) p=0.209	-13% (-30, 4) p=0.129	+11% (4, 18) **p=0.002**	+14% (-4, 39) p=0.126	-2% (-5, 1) p=0.224	-25% (-90, 40) p=0.465	+16% (-2, 35) p=0.090	+7% (-33, 47) p=0.717	-2% (-13, 8) p=0.696

Abbreviations: Conc., concentrated; CI, confidence interval; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated

Supplementary Table 7: Non-head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on BALF and BWF versus the MRS stratified by previous TB status. Trends were similar to the head-to-head analysis (**Table 3**). Data are %, 95% CI, and n/N.

	All patients (n=320)				**No previous TB (n=204)**				**Previous TB (n=116)**			
	Sensitivity	**Sensitivity**	**Sensitivity**	**Sensitivity**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	58 (41, 74) 22/38	95 (91, 97) 267/282	59 (42, 75) 22/37	94 (91, 97) 267/283	61 (42, 78) 19/31	94 (90, 97) 163/173	66 (46, 82) 19/29	93 (88, 96) 163/175	43 (10, 82) 3/7 p=0.372	95 (90, 98) 104/109 p=0.664	38 (9, 76) 3/8 p=0.153	96 (91, 99) 104/108 p=0.265
	n=324				**n=205**				**n=119**			
Unconc. Ultra (study)	76 (60, 88) 31/41 p=0.094‡	71 (65, 76) 200/283 **p<0.001**‡	27 (19, 36) 31/114 **p<0.001**‡	95 (91, 98) 200/210 p=0.661‡	82 (65, 93) 27/33 p=0.068‡	70 (62, 77) 120/172 **p<0.001**‡	34 (24, 46) 27/79 **p=0.004**‡	95 (90, 98) 120/126 p=0.450‡	50 (16, 84) 4/8 p=0.782‡ **p=0.060**	72 (63, 80) 80/111 **p<0.001**‡ p=0.678	11 (3, 27) 4/35 p=0.072‡ **p=0.012**	95 (88, 99) 80/84 p=0.716‡ p>0.999
	n=289				**n=185**				**n=104**			
Conc. Ultra (study)	89 (74, 97) 32/36 p=0.132± **p=0.003**§	77 (71, 82) 195/253 p=0.093± **p<0.001**§	36 (26, 46) 32/90 p=0.199± **p=0.0113**§	98 (95, 99) 195/199 **p=0.048**± p=0.126§	97 (82, 100) 28/29 p=0.067± **p=0.001**§	79 (72, 85) 123/156 p=0.061± **p<0.001**§	46 (33, 59) 28/61 p=0.159± p=0.082 §	99 (96, 100) 123/124 p=0.058± p=0.012§	57 (18, 90) 4/7 p=0.782± p=0.593§ **p=0.003**	74 (64, 83) 72/97 p=0.727± **p<0.001**§ p=0.395	14 (4, 32) 4/29 p=0.776± p=0.130§ **p=0.003**	96 (89, 99) 72/75 p=0.815± p=0.918§ p=0.120
	n=323				**n=206**				**n=117**			
uIFN-γ (rule-in cut-off 4.2 pg/ml)	23 (11, 38) 9/40 **p<0.001**¥	95 (92, 97) 269/283 **p<0.001**¥	39 (20, 61) 9/23 p=0.251¥	90 (86, 93) 269/300 **p=0.023**¥	25 (11, 43) 8/32 **p<0.001**¥	96 (92, 98) 167/174 **p<0.001**¥	53 (27, 79) 8/15 p=0.159¥	87 (82, 92) 167/191 **p=0.020**¥	13 (0, 53) 1/8 p=0.106¥ p=0.449	94 (87, 97) 102/109 **p<0.001**¥ p=0.365	13 (0, 53) 1/8 p=0.932¥ p=0.056	94 (87, 97) 102/109 p=0.622¥ p=0.093

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ± unconc. Ultra (study) vs. conc. Ultra (study), §Xpert vs. conc. Ultra (study), Ψ unconc. Ultra (study) vs. uIFN-γ in patients with the same fluid type (overall, bronchoalveolar lavage, bronchial wash).

Within row p-values: No previous TB vs. previous TB.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; CI, confidence interval; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; unconc., unconcentrated; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 8: Head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on BALF and BWF versus the MRS, eMRS and CRS. The relative performance of the MRS was similar when compared to both the eMRS and CRS. Data are %, 95% CI, and n/N.

	MRS (n=276)				**eMRS (n=276)**				**CRS (n=276)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	59 (41, 75) 20/34	96 (93, 98) 232/242	67 (47, 83) 20/30	94 (91, 97) 232/246	49 (33, 65) 21/43 p=0.383*	96 (93, 98) 224/233 p=0.881*	70 (51, 85) 21/30 p=0.781*	91 (87, 94) 224/246 p=0.166*	49 (33, 65) 21/43 p>0.999§ p=0.383†	96 (93, 98) 224/233 p>0.999§ p=0.881†	70 (51, 85) 21/30 p>0.999§ p=0.781†	91 (87, 94) 224/246 p>0.999§ p=0.166†
Unconc. Ultra (study)	79 (61, 91) 26/33 p=0.078‡	72 (66, 77) 174/242 **p<0.001‡**	28 (20, 39) 27/95 **p<0.001‡**	96 (92, 98) 174/181 p=0.389‡	77 (61, 88) 33/43 **p=0.007‡** p=0.832*	73 (67, 79) 171/233 **p<0.001‡** p=0.716*	35 (25, 45) 33/95 **p=0.001‡** p=0.349*	94 (90, 97) 171/181 p=0.185‡ p=0.456*	77 (61, 88) 33/43 **p=0.007‡** p>0.999§ p=0.832†	73 (67, 79) 171/233 **p<0.001‡** p>0.999§ p=0.716†	35 (25, 45) 33/95 **p=0.001‡** p>0.999§ p=0.349†	94 (90, 97) 171/181 p=0.185‡ p>0.999§ p=0.456†
Conc. Ultra (study)	91 (76, 98) 31/34 p=0.155±	77 (71, 82) 187/242 p=0.175±	36 (26, 47) 31/86 p=0.272±	98 (95, 100) 187/190 p=0.174±	81 (67, 92) 35/43 p=0.596± p=0.223*	78 (72, 83) 182/233 p=0.235± p=0.826*	41 (30, 52) 35/86 p=0.408± p=0.531*	96 (92, 98) 182/190 p=0.556± p=0.126*	81 (67, 92) 35/43 p=0.596± p>0.999§ p=0.223†	78 (72, 83) 182/233 p=0.235± p>0.999§ p=0.826†	41 (30, 52) 35/86 p=0.408± p>0.999§ p=0.531†	96 (92, 98) 182/190 p=0.556± p>0.999§ p=0.126†
uIFN-γ (rule-out cut-off 5.1 pg/ml)	18 (7, 35) 6/34 **p<0.001¥**	96 (93, 98) 232/242 **p<0.001¥**	38 (15, 65) 6/16 p=0.462¥	89 (85, 93) 232/260 **p=0.008¥**	16 (7, 31) 7/43 **p<0.001¥** p=0.874*	96 (93, 98) 224/233 **p<0.001¥** p=0.881*	44 (20, 70) 7/16 p=0.487¥ p=0.719*	86 (81, 90) 224/260 **p=0.005¥** p=0.286*	16 (7, 31) 7/43 **p<0.001¥** p>0.999§ p=0.874†	96 (93, 98) 224/233 **p<0.001¥** p>0.999§ p=0.881†	44 (20, 70) 7/16 p=0.487¥ p>0.999§ p=0.719†	86 (81, 90) 224/260 **p=0.005¥** p>0.999§ p=0.286†

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ± unconc. Ultra (study) vs. conc. Ultra (study), ¥unconc. Ultra (study) vs. uIFN-γ in patients using the same reference standard (MRS, eMRS, CRS).

Within row p-values: *MRS vs. eMRS, §eMRS vs. CRS, †MRS vs. CRS.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; unconc., unconcentrated; uIFN-γ, unstimulated interferon gamma; Ultra; Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 9: Non-head-to-head diagnostic accuracy analyses of Xpert, Ultra (unconc. and conc.) and uIFN-γ on BALF and BWF versus the MRS, eMRS and CRS. Trends were similar to non-head-to-head analyses (**Supplementary Table 8**). Data are %, 95% CI, and n/N.

	MRS (n=320)				**eMRS (n=338)**				**CRS (n=346)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Xpert (programmatic conc. and study unconc. pooled)	58 (41, 74) 22/38	95 (91, 97) 267/282	59 (42, 75) 22/37	94 (91, 97) 267/283	48 (34, 62) 25/52 p=0.357*	93 (90, 96) 267/286 p=0.506*	57 (41, 72) 25/44 p=0.810*	91 (87, 94) 267/294 p=0.107*	47 (33, 61) 25/53 p=0.926§ p=0.313†	93 (90, 96) 273/293 p=0.930§ p=0.450†	56 (40, 70) 25/45 p=0.905§ p=0.722†	91 (87, 94) 273/301 p=0.960§ p=0.095†
	n=324				**n=341**				**n=349**			
Unconc. Ultra (study)	76 (60, 88) 31/41 p=0.094‡	71 (65, 76) 200/283 **p<0.001**‡	27 (19, 36) 31/114 **p<0.001**‡	95 (91, 98) 200/210 p=0.661‡	73 (59, 84) 40/55 **p=0.009**‡ p=0.750*	71 (65, 76) 202/286 **p<0.001**‡ p=0.991*	32 (24, 41) 40/124 **p=0.004**‡ p=0.394*	93 (89, 96) 202/217 p=0.356‡ p=0.344*	71 (58, 83) 40/56 **p=0.010**‡ p=0.879§ p=0.646†	71 (65, 76) 207/293 **p<0.001**‡ p=0.996§ p=0.995†	32 (24, 41) 40/126 **p=0.005**‡ p=0.931§ p=0.440†	93 (89, 96) 207/223 p=0.385‡ p=0.914§ p=0.291†
	n=289				**n=306**				**n=314**			
Conc. Ultra (study)	89 (74, 97) 32/36 p=0.132±	77 (71, 82) 195/253 p=0.093±	36 (26, 46) 32/90 p=0.199±	98 (95, 99) 195/199 p=0.126±	81 (67, 91) 39/48 p=0.307± p=0.338*	77 (71, 82) 198/258 p=0.107± p=0.929*	39 (30, 50) 39/99 p=0.268± p=0.586*	96 (92, 98) 198/207 p=0.253± p=0.181*	81 (67, 91) 39/48 p=0.243± p>0.999§ p=0.338†	76 (71, 81) 203/266 p=0.130± p=0.908§ p=0.838†	38 (29, 48) 39/102 p=0.306± p=0.866§ p=0.701†	96 (92, 98) 203/212 p=0.189± p=0.959§ p=0.196†
	n=323				**n=340**				**n=348**			
uIFN-γ (rule-out cut-off 5.1 pg/ml)	23 (11, 38) 9/40 **p<0.001**¥	95 (92, 97) 269/283 **p<0.001**¥	39 (20, 61) 9/23 p=0.251¥	90 (86, 93) 269/300 **p=0.023**¥	20 (11, 34) 11/54 **p<0.001**¥ p=0.803*	95 (92, 97) 272/286 **p<0.001**¥ p=0.977*	44 (24, 65) 11/25 p=0.259¥ p=0.733*	86 (82, 90) 272/315 **p=0.014**¥ p=0.206*	20 (10, 33) 11/55 **p<0.001**¥ p=0.962§ p=0.768†	95 (92, 97) 279/293 **p<0.001**¥ p=0.948§ p=0.925†	44 (24, 65) 11/25 p=0.237¥ p>0.999§ p=0.733†	86 (82, 90) 279/323 **p=0.018**¥ p=0.992§ p=0.208†

Within column p-values: ‡Xpert vs. unconc. Ultra (study), ± unconc. Ultra (study) vs. conc. Ultra (study), ¥unconc. Ultra (study) vs. uIFN-γ in patients using the same reference standard (MRS, eMRS, CRS).

Within row p-values: *MRS vs. eMRS, §eMRS vs. CRS, †MRS vs. CRS.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; unconc., unconcentrated; uIFN-γ, unstimulated interferon gamma; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 10: All non-actionable unconcentrated Xpert, unconcentrated and concentrated Ultra results (programmatic Xpert concentrated results were not available). If a sample had sufficient volume, a test was repeated once. The most common Xpert error was 5011: Signal loss detected in the amplification curve for analyte SPC and the most common error for concentrated Ultra was 2008: Syringe pressure reading the protocol limit. Data are % and n/N.

Test	Non-actionable rate	Patient ID	BALF or BWF	Specimen grade	First result	Error code	Second result	Error code/reason test was not repeated	MRS status
Unconc. Xpert (study)	3% (10/299)	BRO070	BALF	Opaque	Error	5011	Positive	-	Positive
		BRO087	BALF	Opaque	Error	5011	Positive	-	Negative
		BRO097	BALF	Opaque	Error	5011	Negative	-	Negative
		BRO098	BWF	Opaque	Error	5011	Positive	-	Negative
		BRO134	BWF	Opaque	Error	5007	Negative	-	Negative
		BRO184	BWF	Opaque	Error	5011	Negative	-	Negative
		BRO257	BWF	Clear	No result	2037	Positive	-	Positive
		BRO259	BALF	Clear	Error	5011	Negative	-	Negative
		BRO302	BWF	Clear	Error	5011	Not done	No more fluid	Positive
		BRO304	BWF	Clear	Error	5011	Not done	No more fluid	Positive
8/8 non-actionable results resolved upon repeat testing (4 Xpert-positive, 4 Xpert-negative)									
Unconc. Ultra (study)	1% (2/354) **p=0.008**‡	BRO017	BALF	Clear	Error	5011	Negative	-	Negative
		BRO026	BALF	Clear	Error	5017	Positive	-	Negative
2/2 non-actionable results resolved upon repeat testing (1 Ultra-positive, 5 Ultra-negative)									
Conc. Ultra (study)	4% (12/335) **p=0.005**± p=0.870¥	BRO031	BALF	Opaque	No result	2037	Negative	-	Negative
		BRO044	BWF	Clear	Invalid	5004	Not done	No more fluid	Negative
		BRO051	BALF	Clear	Error	5011	Negative	-	Negative
		BRO171	BWF	Opaque	Error	5011	Not done	No more fluid	Negative
		BRO202	BALF	Clear	Error	2008	Not done	No more fluid	Negative
		BRO215	BWF	Clear	Error	2008	Positive	-	Negative
		BRO220	BALF	Clear	Error	2008	Negative	-	Negative
		BRO251	BALF	Clear	Error	2008	Negative	-	Negative

		BRO302	BWF	Clear	Error	2008	Not done	No more fluid	Positive
		BRO306	BWF	Clear	Error	2008	Not done	No more fluid	Negative
		BRO331	BWF	Clear	Error	2008	Negative	-	Negative
		BRO339	BWF	Opaque	Error	2008	Not done	No more fluid	Negative
6/6 non-actionable results resolved upon repeat testing (1 Ultra-positive, 5 Ultra-negative)									

Within column p-values: ‡ Study unconcentrated Xpert vs. study unconcentrated Ultra, ± study unconcentrated Ultra vs. study concentrated Ultra, ¥ study unconcentrated Xpert vs. study concentrated Ultra

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; unconc. unconcentrated.

Error code definitions:
2037: The cartridge integrity test failed.
2008: Syringe pressure reading the protocol limit.
5004: Failed to verify valid amplification curve for analyte *rpo*B.
5011: Signal loss detected in the amplification curve for analyte SPC.
5017: SPC probe check failed.

Supplementary Table 11: Rifampicin (RIF) results for Xpert (programmatic and study) and study Ultra (unconc. and conc.) on BALF and BWF culture isolates that had programmatic RIF-susceptibility testing (MTBDR*plus*) done. Conc. Ultra incorrectly classified one RIF-susceptible case as RIF-resistant. Data %, 95% CI and n/N.

MTBDR*plus* result	**Xpert-positive***	**Unconc. Ultra-positive (study)**	**Conc. Ultra-positive (study)**
RIF-resistant (1/11)	**RIF-resistant**		
	100% (1/1)	100% (1/1)	100% (1/1)
	RIF-susceptible		
	0% (0/1)	0% (0/1)	0% (0/1)
	RIF-indeterminate		
	0% (0/1)	0% (0/1)	0% (0/1)
RIF-susceptible (10/11)	**RIF-resistant**		
	0% (0/1)	0% (0/6)	14% (1/7)
	RIF-susceptible		
	100% (1/1)	50% (3/6)	43% (3/7)
	RIF-indeterminate		
	0% (0/1)	50% (3/6)	43% (3/7)

*Xpert included study unconcentrated and programmatic concentrated results.

Missing data: Study conc. Ultras not done=2.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; MTBDR*plus*, GenoType MTBDR*plus*; RIF, rifampicin; unconc., unconcentrated.

Supplementary Table 12: Concordance of study Conc. and unconc. Ultra vs. programmatic conc. Ultra and study conc. Ultra vs. study unconc. Ultra on BALF or BWF in patients with both tests done. Study conc. Ultras had more positive results than programmatic conc. Ultras and unconc. Ultras had more positive results than programmatic conc. Ultras.

A		**Conc. Ultra (study)**		
		Positive	Negative	Total
Conc. Ultra (programmatic)	Positive	40	6	46
	Negative	35	159	194
	Total	75	165	240
	Δ Conc. Ultra (study) vs. conc. Ultra (programmatic)	+12% (95% confidence interval; CI: 7, 18) **p<0.001**		
B		**Unconc. Ultra (study)**		
		Positive	Negative	Positive
Conc. Ultra (programmatic)	Positive	49	13	62
	Negative	51	158	209
	Total	100	171	271
	Δ Unonc. Ultra (study) vs. conc. Ultra (programmatic)	+14% (95% confidence interval; CI: 8, 20) **p<0.001**		
C		**Conc. Ultra (study)**		
		Positive	Negative	Positive
Unconc. Ultra (study)	Positive	70	43	113
	Negative	36	173	209
	Total	106	216	322
	Δ Unonc. Ultra (study) vs. conc. Ultra (study)	+2% (95% confidence interval; CI: -4, 9) p=0.431		

Non-actionable rates of programmatic concentrated Ultras are 1 (was study concentrated-negative). Non-actionable rates of study concentrated Ultras were 13 (6 were programmatic concentrated-positive and 7 were programmatic concentrated-negative-all of these had actionable study unconcentrated Ultra results). Non-actionable rates of study unconcentrated Ultras are 1 (study concentrated-negative). Non-actionable rates of study concentrated Ultras are 14 (8 were study unconcentrated-positive and 6 were study unconcentrated-negative).

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; CI, confidence interval; conc., concentrated; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated.

Supplementary Table 13: Non-head-to-head and head-to-head diagnostic accuracy analyses of Ultra (unconc. or conc.) on urine and unconc. Ultra on BALF and BWF versus the MRS by HIV status. Urine has low utility for diagnosing TB. Data are %, 95% CI, and n/N.

	Non-head-to-head											
	All patients				**HIV-negative**				**HIV-positive**			
	n=53				**n=43/53 (81)**				**n=10/53 (19)**			
	Sensitivity	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Ultra on urine (unconc. or conc.)	0 (0, 46) 0/6	91 (80, 98) 43/47	0 (0, 60) 0/4	88 (75, 95) 43/49	0 (0, 46) 0/6	95 (82, 99) 35/37	0 (0, 84) 0/2	88 (74, 96) 36/41	0	80 (44, 97) 8/10 p=0.142*	0 (0, 84) 0/2 p>0.999*	100 (63, 100) 8/8 p=0.297*
	n=324				**n=273/324 (84)**				**n=51/324 (16)**			
Unconc. Ultra on BALF or BWF (study)	76 (60, 88) 31/41 **p<0.001‡**	71 (65, 76) 200/283 **p=0.003‡**	27 (19, 36) 31/114 p=0.225‡	95 (91, 98) 200/210 **p=0.050‡**	81 (64, 92) 29/36 **p<0.001‡**	71 (65, 77) 169/237 **p=0.003 ‡**	30 (21, 40) 29/97 p=0.358‡	96 (92, 98) 169/176 **p=0.038‡**	40 (5, 85) 2/5 **p=0.048***	67 (52, 80) 31/46 p=0.432‡ p=0.593*	12 (1, 36) 2/17 p=0.608‡ p=0.121*	91 (76, 98) 31/34 p=0.383‡ p=0.225*
	Head-to-head											
	n=52				**n=42/52 (81)**				**n=10/52 (19)**			
Ultra on urine (unconc. or conc.)	0 (0, 46) 0/6	93 (82, 99) 43/46	0 (0, 71) 0/3	88 (75, 95) 43/49	0 (0, 46) 0/6	97 (85, 100) 35/36	0 (0, 97) 0/1	85 (71, 94) 35/41	0	80 (44, 97) 8/10 p=0.051*	0 (0, 84) 0/2 p=0.999*	100 (63, 100) 8/8 p=0.248*
Unconc. Ultra on BALF or BWF (study)	67 (22, 96) 4/6 **p=0.014‡**	72 (57, 84) 33/46 **p=0.006‡**	24 (7, 50) 4/17 p=0.348‡	94 (81, 99) 33/35 p=0.315‡	67 (22, 96) 4/6 **p=0.014‡**	78 (61, 90) 28/36 **p=0.013‡**	33 (10, 65) 4/12 p=0.488‡	93 (78, 99) 28/30 p=0.294‡	0	50 (19, 81) 5/10 p=0.160‡ p=0.084*	0 (0, 52) 0/5 p>0.999‡ p=0.140*	100 (48, 100) 5/5 p>0.999‡ p=0.552*

Within column p-values: ‡ Urine-Ultra vs. Ultra on BALF or BWF in patients of the same HIV status (overall, negative, positive).

Within row p-values: *HIV-negative vs. HIV-positive.

Abbreviations: BALF, bronchoalveolar lavage fluid; BWF, bronchial wash fluid; conc., concentrated; CI, confidence interval; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; NPV, negative predictive value; PPV, positive predictive value; TB, tuberculosis; unconc., unconcentrated; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Supplementary Table 14: Per patient information for study unconcentrated Ultra-positive patients (on BF) that were MRS-negative, with previous TB status, Ultra semi-quantitation category, programmatic Xpert or Ultra results, alternate reference standard result and treatment initiation status after at least 12-weeks follow-up. Approximately 7 in 10 MRS-negative Ultra-positives were trace-positive and 1 in 4 started empirical treatment. Data are n/N (%).

Patient ID	Previous TB	Ultra semi-quantitation category	Programmatic Xpert or Ultra result*	Positive by eMRS and/or CRS	Treatment initiated after 12-week follow up
BRO006	Yes	Very low	Negative	Negative	No
BRO008	Yes	Very low	Not Done	Negative	No
BRO016	Yes	Trace	Negative	Negative	Unsuccessful follow up after 3 calls
BRO019	Yes	Low	Positive	Negative	Yes
BRO026	Yes	Trace	Negative	Negative	No
BRO040	Yes	Very low	Negative	Negative	Unsuccessful follow up after 3 calls
BRO054	No	Very low	Negative	Negative	No
BRO057	No	Trace	Positive	Negative	Unsuccessful follow up after 3 calls
BRO066	Yes	Trace	Negative	Negative	No
BRO067	No	Trace	Negative	Negative	No
BRO071	Yes	Very low	Negative	Negative	No
BRO073	Yes	Trace	Negative	Negative	No
BRO077	No	Trace	Negative	Negative	Yes
BRO079	Yes	Trace	Positive	Negative	Yes
BRO083	No	Trace	Negative	Negative	No
BRO085	Yes	Trace	Negative	Negative	No
BRO086	No	Trace	Negative	Negative	No
BRO087	No	Trace	Positive	Negative	No
BRO092	No	Trace	Negative	Negative	No
BRO093	No	Trace	Negative	Negative	No
BRO096	Yes	Trace	Positive	Negative	Yes
BRO098	No	Very low	Positive	Negative	Yes
BRO102	Yes	Trace	Positive	Negative	Yes
BRO104	No	Very low	Negative	Negative	No
BRO116	No	Trace	Positive	Negative	Yes
BRO117	Yes	Trace	Negative	Negative	No
BRO118	Yes	Low	Positive	Negative	Yes
BRO125	Yes	Trace	Negative	Negative	No
BRO127	Yes	Trace	Negative	Negative	No
BRO152	No	Very low	Positive	Negative	No
BRO154	No	Trace	Negative	Negative	No
BRO169	No	Very low	Negative	Negative	No
BRO174	No	Very low	Negative	Negative	Yes
BRO176	No	Trace	Not done	Negative	Yes

BRO182	No	Low	Negative	Negative	No
BRO185	No	Trace	Negative	Negative	No
BRO186	No	Very low	Negative	Negative	No
BRO187	No	Very low	Negative	Negative	No
BRO189	No	Trace	Not done	Negative	No
BRO191	No	Trace	Negative	Negative	No
BRO195	No	Trace	Negative	Negative	Yes
BRO202	No	Very low	Positive	Negative	Yes
BRO209	No	Trace	Negative	Negative	No
BRO212	No	Trace	Negative	Negative	No
BRO214	No	Trace	Positive	Negative	No
BRO215	Yes	Trace	Negative	Negative	No
BRO219	No	Trace	Negative	Negative	No
BRO220	No	Trace	Negative	Negative	No
BRO223	No	Trace	Negative	Negative	No
BRO225	No	Trace	Negative	Negative	Yes
BRO233	No	Trace	Negative	Negative	No
BRO235	Yes	Trace	Negative	Negative	No
BRO247	Yes	Very low	Negative	Negative	No
BRO254	No	Trace	Negative	Negative	No
BRO255	No	Very low	Not done	Negative	No
BRO259	No	Trace	Negative	Negative	Yes
BRO260	No	Trace	Negative	Negative	No
BRO261	No	Low	Negative	Negative	No
BRO269	Yes	Trace	Positive	Negative	Yes
BRO273	No	Trace	Negative	Negative	No
BRO274	No	Very low	Negative	Negative	No
BRO276	No	Trace	Negative	Negative	No
BRO277	No	Very low	Negative	Negative	No
BRO280	Yes	Very low	Positive	Negative	Yes
BRO285	No	Trace	Negative	Negative	No
BRO288	No	Low	Negative	Negative	No
BRO290	No	Trace	Not done	Negative	Yes
BRO291	Yes	Very low	Negative	Negative	No
BRO294	Yes	Very low	Positive	Negative	No
BRO295	No	Trace	Negative	Negative	No
BRO296	Yes	Trace	Negative	Negative	No
BRO298	Yes	Trace	Negative	Negative	No
BRO306	Yes	Very low	Positive	Negative	Yes
BRO318	Yes	Trace	Not done	Negative	No
BRO319	No	Trace	Not done	Negative	Unsuccessful follow up after 3 calls
BRO321	No	Trace	Positive	Negative	Unsuccessful follow up after 3 calls
BRO323	Yes	Trace	Positive	Negative	No
BRO332	No	Trace	Negative	Negative	No

BRO333	No	Trace	Negative	Negative	No
BRO337	No	Trace	Negative	Negative	Unsuccessful follow up after 3 calls
BRO354	No	Trace	Positive	Negative	No
BRO359	Yes	Trace	Positive	Negative	Unsuccessful follow up after 3 calls
BRO360	Yes	Trace	Negative	Negative	No
Overall	31/83 (37)	trace: 58/83 (70) very low: 19/83 (23) low: 4/83 (5) medium: 2/83 (2)	Negative: 56/83 (68) Positive: 20/83 (24) Non-actionable: 7/83 (8)	Both: 7/83 (8) Negative: 76/83 (92)	Yes:18/76 (24) No: 58/76 (76)

*Not done because of the programmatic algorithm in **Figure 1**.

Abbreviations: BF, bronchial fluid; CRS, composite reference standard; eMRS, extended microbiological reference standard; MRS, microbiological reference standard; Ultra, Xpert MTB/RIF Ultra; Xpert, Xpert MTB/RIF.

Missing data: Treatment initiated, 7.

Supplementary Table 15: Diagnostic yield (proportion of people with at least one positive confirmatory test result detected by a test) of programmatic and study microbiological tests. BF Study Ultra had the highest diagnostic yield followed by Xpert and culture on BF. No significant difference in diagnostic yield was observed when tests were stratified by HIV status. Data are n/N (%).

Test	Diagnostic yield			
	In patients who did or did not have the test attempted	**In people without HIV in patients who did or did not have the test attempted**	**In PLHIV in patients who did or did not have the test attempted**	**In patients who had the test attempted**
BF				
Study Ultra (unconc. and conc.)	165/185 (89)	136/153 (89)	28/31 (90) p=0.815	165/354 (47)
*Xpert (unconc. and conc.)	48/185 (26)	41/153 (27)	6/31 (19) p=0.386	48/350 (14)
Culture	41/185 (22)	36/153 (24)	5/31 (16) p=0.367	41/325 (13)
Lung biopsy				
Culture	0/185 (0)	0/153 (0)	0/31 (0) p>0.999	0/8 (0)
Non-site-of-disease fluid				
Smear microscopy	4/185 (2)	4/153 (3)	0/31 (0) p=0.363	4/4 (100)
Culture	15/185 (8)	11/153 (7)	4/31 (13) p=0.289	15/131 (11)
Programmatic Ultra	7/185 (4)	6/153 (4)	1/31 (3) p=0.854	13/109 (6)
Programmatic Xpert	1/185 (1)	1/153 (1)	0/31 (0) p=0.652	1/34 (3)
Urine				
Ultra (unconc. and conc.)	4/185 (2)	2/153 (1)	2/31 (6) p=0.073	4/56 (7)

*Includes programmatic concentrated Xpert and study unconcentrated Xpert (latter done by the study if former not done programmatically- only BF was used).

Abbreviations: BF, bronchial fluid; conc., concentrated; MGIT960 culture, Mycobacteria Growth Indicator Tube 960; Ultra, Xpert MTB/RIF Ultra; unconc., unconcentrated; Xpert, Xpert MTB/RIF.

Appendix V

Extract from used Xpert MTB/RIF Ultra cartridges is useful for accurate second-line drug-resistant tuberculosis diagnosis with minimal *rpoB*-amplicon cross-contamination risk

Publication status: published

Venter, R.[*], **Minnies, S[*].**, Derendinger, B., Tshivhula, H., De Vos, M., Dolby, T., Ruiters, A., Warren, R.M. and Theron, G., 2020. Extract from used Xpert MTB/RIF Ultra cartridges is useful for accurate second-line drug-resistant tuberculosis diagnosis with minimal rpoB-amplicon cross-contamination risk. *Scientific reports*, *10*(1), pp.1-10.

DOI: 10.1038/s41598-020-59164-3

[*]***Contributed equally***

SCIENTIFIC REPORTS
nature research

OPEN

Extract from used Xpert MTB/RIF Ultra cartridges is useful for accurate second-line drug-resistant tuberculosis diagnosis with minimal *rpoB*-amplicon cross-contamination risk

Rouxjeane Venter[1,3], Stephanie Minnies[1,3], Brigitta Derendinger[1], Happy Tshivhula[1], Margaretha de Vos[1], Tania Dolby[2], Ashley Ruiters[2], Robin M. Warren[1] & Grant Theron[1*]

Xpert MTB/RIF Ultra (Ultra) detects *Mycobacterium tuberculosis* and rifampicin resistance. Follow-on drug susceptibility testing (DST) requires additional sputum. Extract from the diamond-shaped chamber of the cartridge (dCE) of Ultra's predecessor, Xpert MTB/RIF (Xpert), is useful for MTBDR*sl*-based DST but this is unexplored with Ultra. Furthermore, whether CE from non-diamond compartments is useful, the performance of FluoroType MTBDR (FT) on CE, and *rpoB* cross-contamination risk associated with the extraction procedure are unknown. We tested MTBDR*sl*, MTBDR*plus*, and FT on CEs from chambers from cartridges (Ultra, Xpert) tested on bacilli dilution series. MTBDR*sl* on Ultra dCE on TB-positive sputa (n = 40) was also evaluated and, separately, *rpoB* amplicon cross-contamination risk . MTBDR*sl* on Ultra dCE from dilutions $\geq 10^3$ CFU/ml (C_{Tmin} <25, >"low semi-quantitation") detected fluoroquinolone (FQ) and second-line injectable (SLID) susceptibility and resistance correctly (some SLIDs-indeterminate). At the same threshold (at which ~85% of Ultra-positives in our setting would be eligible), 35/35 (100%) FQ and 34/35 (97%) SLID results from Ultra dCE were concordant with sputa results. Tests on other chambers were unfeasible. No tubes open during 20 batched extractions had FT-detected *rpoB* cross-contamination. False-positive Ultra *rpoB* results was observed when dCE dilutions $\leq 10^{-3}$ were re-tested. MTBDR*sl* on Ultra dCE is concordant with isolate results. *rpoB* amplicon cross-contamination is unlikely. These data mitigate additional specimen collection for second-line DST and cross-contamination concerns.

Drug-resistant tuberculosis (TB) remains a global threat[1]. Of 10 million estimated incidence cases reported in 2017, 588 000 were rifampicin-resistant[2]. Of these ~458 000 were multidrug-resistant (MDR). Despite the improved roll-out of rifampicin-resistance testing, many patients are not diagnosed appropriately or started on effective treatment, resulting in huge TB care cascade gaps[3,4]. For example, in South Africa, 84% of patients with drug-resistant TB have access to rifampicin-susceptibility testing, but only 47% of these are started on likely effective treatment[4]. Similarly, in India, only 41% of the MDR-TB burden was diagnosed in 2013 and, of these, just 32% started on treatment[5]. Innovative approaches are needed to ensure more patients receive comprehensive drug susceptibility testing (DST).

Previous work showed that mycobacterial genomic DNA can be recovered from the rear diamond-shaped chamber of used Xpert MTB/RIF (Xpert) cartridges after the test is complete. This diamond cartridge extract (dCE) is useful for downstream testing with the MTBDR*sl* line probe assay (LPA) (Hain Lifescience, Germany), the only World Health Organization (WHO)-endorsed molecular test for second-line drug resistance, and

[1]DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, SA MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. [2]National Health Laboratory Services, Cape Town, South Africa. [3]These authors contributed equally: Rouxjeane Venter and Stephanie Minnies. *email: gtheron@sun.ac.za

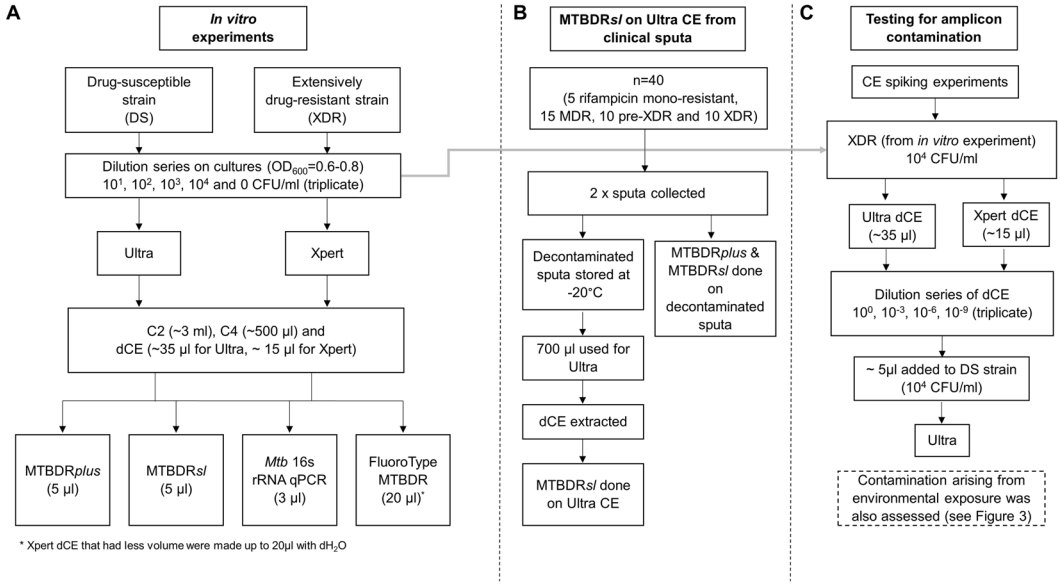


Figure 1. Study flow diagrams for the (**A**) *in vitro* experiment, (**B**) MTBDR*sl* on Ultra CE from clinical sputa experiment, and the (**C**) evaluation of *rpoB* amplicon cross-contamination risk experiment.

spoligotyping[6], a method useful for monitoring the molecular epidemiology of TB outbreaks. This additional testing does not require extra specimen collection nor additional downstream DNA extraction, both of which can exacerbate patient loss within the diagnostic care cascade.

As Xpert is a real-time PCR that generates quantitative information, a cycle threshold value (C_T <24) was identified at which downstream dCE testing using MTBDR*sl* was successful and fully concordant with MTBDR*sl* results on matching isolates[7]. However, Xpert dCE was not useful for first-line DST using the WHO-endorsed MTBDR*plus* assay, likely due to interference from large numbers of Xpert *rpoB* amplicons. In addition to the dCE approach, others[8,9] have shown it is possible to test leftover specimen-sample reagent mix remaining after Xpert, however, remnant volume is not always present and DNA extraction and downstream clean-up might still be needed.

Xpert MTB/RIF Ultra (Ultra) recently superseded Xpert as WHO-endorsed frontline molecular test-of-choice for TB and rifampicin resistance[10]. Compared to Xpert, Ultra has higher sensitivity in paucibacillary samples, however, specificity is overall lower[11–13]. Ultra is a different assay compared to Xpert and it is not necessarily given that the extract approach would be feasible on Ultra dCE. We aimed to confirm that Ultra dCE would be useful for second-line DST. Furthermore, we asked if extract from other chambers within the cartridge other than the diamond (i.e., chambers that are likely *rpoB* amplicon-free), may contain DNA. We quantified this DNA using a *Mycobacterium tuberculosis* complex 16S rRNA real time qPCR and evaluated whether this DNA was useful for first-line DST using the FluoroType MTBDR (FT) (Hain Lifescience, Germany) assay[14,15]. A test such as FT could, for example, be used to check for isoniazid mono-resistance or confirm Ultra rifampicin-resistance results.

Lastly, as the cartridge extraction (CE) procedure involves aspirating fluid rich in *rpoB* amplicons, it may represent a source of cross-contamination. We sought to evaluate this risk, both under a prolonged exposure scenario (where collection tubes were purposely exposed during extended batch extractions) and an absolute worst-case scenario (directly adding dCE to a sample later tested by Ultra). Showing that the extracted cartridge approach in Ultra is compatible with MTBDR*sl* and represents minimal *rpoB* amplicon cross-contamination risk would increase the likelihood of implementation, especially as Xpert is in the process of being phased out in lieu of Ultra. In turn, this could reduce both sputum collection requirements for complete DST and time-to-effective-treatment initiation.

Methods

Ethics statement. Methods and protocols were carried out in accordance with relevant guidelines and regulations. The study was approved by the Health Research Ethics Committee of Stellenbosch University (N09/11/296) and the City of Cape Town (10570). Permission was granted to use anonymised residual specimens collected during routine diagnostic practice and thus patient informed consent was waived.

Ultra and Xpert on dilution series of *Mycobacterium tuberculosis* strains. Culturing of genotypically-confirmed drug-susceptible (DS-TB) and extensively-drug resistant (XDR) *M. tuberculosis* isolates were done in a Biosafety Level (BSL) 3 laboratory to an OD_{600} of 0.6–0.8 (Fig. 1A). A triplicate tenfold dilution series from three separate cultures [10^0–10^4 colony forming units (CFU)/ml] was prepared in phosphate buffer (33 mM Na_2HPO_4, 33 mM KH_2PO_4; pH 6.8) with 0.025% Tween80 (Sigma-Aldrich, United States). Colony counts were done on 7H11 Middlebrook agar (BD Biosciences, United States). A total of 52 dilutions [four dilutions, 10^1–10^4 CFU/ml in triplicate for both strains plus a negative control for each strain; $(4 \times 3 \times 2 + 2) \times 2$] were made up to 1 ml and tested by Ultra (n = 26) or Xpert (n = 26) per the manufacturer's instructions[16,17]. Used positive cartridges were stored prior to extraction at 4 °C for ≤3 days. Crude DNA (heat inactivated for

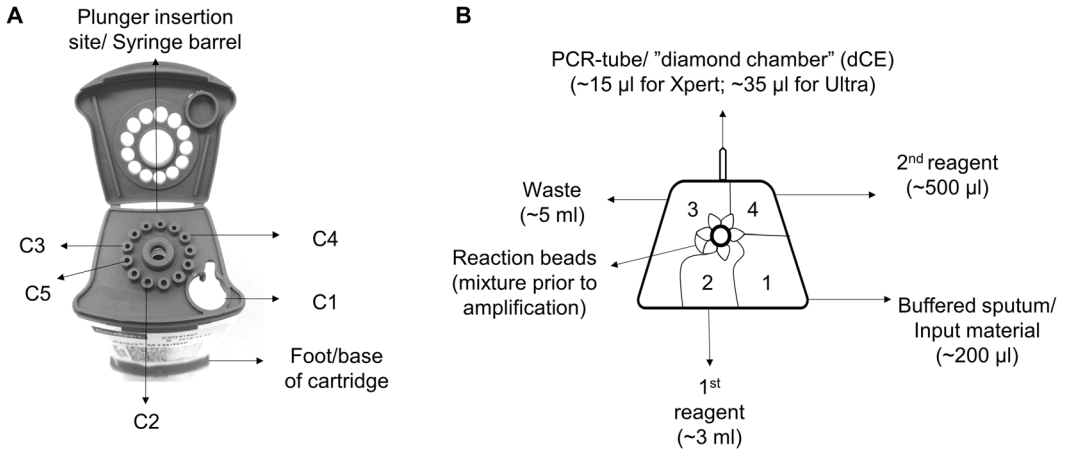


Figure 2. (**A**) Entry points through the lid of the cartridge for access to different cartridge chambers. (**B**) Top-down cross-section of the inside of the cartridge corresponding to the access points.

2 hours at 100 °C) from the same strains served as positive controls for downstream tests (16S rRNA gene qPCR, MTBDR*plus*, MTBDR*sl*, FT).

Ultra on sputum from TB patients. Forty used positive Ultra cartridges done on NALC-NaOH decontaminated sputa from pre-treatment TB patients with known drug resistance [5 rifampicin-mono-resistant, 15 MDR, 10 pre-XDR (resistance to rifampicin, isoniazid and either a fluoroquinolones or a second-line injectable), 10 XDR] were collected from November 2015 to September 2017 and dCEs were extracted as described previously[6] (Fig. 1B). To confirm MTBDR*sl* results from dCEs, MTBDR*sl* was done per the manufacturer's instructions directly on corresponding decontaminated sputa[18,19]. Ultra cartridges were processed in a manner blinded to MTBDR*sl* results.

Recovery of mycobacterial genomic DNA from used Ultra and Xpert cartridges. *Preparation of work space.* BSL2 hood surfaces were sterilised [1% NaOCl (bleach), 70% EtOH, 5 min UV irradiation] before and after each batched extraction. Each cartridge was wiped with 1% bleach and 70% EtOH before and after each extraction.

Description of cartridge design. To investigate the feasibility of testing extract from Ultra and Xpert cartridge chambers, an understanding of their design and inner processes is required. As described previously, each cartridge has a similar design consisting of a foot, valve, body, reaction tube and lid[20,21]. The five internal chambers hold buffers and lyophilised PCR reagents used for sputum homogenisation, washing away debris, and DNA extraction, purification, and amplification[22]. The Xpert and Ultra procedures, including the processes inside the cartridges and the contents of each chamber are described in the supplement. After assay conclusion, the volumes typically remaining in each chamber are ~500 µl for Chamber 1 (C1), ~3 ml for Chamber 2 (C2), ~5 ml for Chamber 3 (C3) and ~500 µl for Chamber 4 (C4) [Chamber 5 (C5) had no volume remaining after test completion].

Diamond chamber extract. dCEs were extracted from all positive cartridges by puncturing the rear chamber with a sterile 29 G × 1/2″ 1 ml insulin syringe (Avacare, South Africa) (Fig. 2A,B) as described previously[6]. The full volume was extracted (~15 µl for Xpert; ~35 µl for Ultra). CEs were stored in microcentrifuge tubes at −20 °C prior to analysis.

Other chambers. Five cartridge chambers (C1, C2, C3, C4, C5) were accessed by inserting a 22 G spinal needle (Becton Dickinson, United States) fixed a 5 ml syringe (Fig. 2A; a pipette may also be used for C1) and the entire volume withdrawn (Fig. 2A,B). C5 had no remaining volume left after Xpert or Ultra test completion. No DNA extraction or purification steps were done for downstream assays.

16s rRNA gene quantitative PCR (qPCR) on cartridge extract. CEs from C1–4 and dCE from Ultra and Xpert done on the serial dilutions were tested (heat extracted crude DNA from matching isolates was used as positive control). For each qPCR, 5 µl iTaq Universal SYBR Green Supermix (Bio-Rad), 0.3 µl (300 nM) of *M. tuberculosis* specific forward (V4 515F) primers, 0.3 µl (300 nM) of *M. tuberculosis* specific reverse (V4 806R) primers (Table S1) and 1.4 µl nuclease-free water was used[23]. 3 µl CE was added and amplification occurred using a Bio-Rad CFX-96. The threshold used to determine if a reaction was excluded from subsequent analyses was defined as a C_q value greater than the average of the triplicate negative controls for that run. Chambers with a C_q less than that average value were considered positive for *M. tuberculosis* complex (MTBC) DNA and used for MTBDR*plus*, MTBDR*sl* and FT.

MTBDR*plus* and MTBDR*sl* line probe assays on cartridge extract. *Diamond chamber extract.* MTBDR*plus* and MTBDR*sl* (both version 2.0) were performed on dCEs from Ultra and Xpert done on the *in vitro* dilution series. For Ultra done on sputa from patients, only MTBDR*sl* was done. 5 µl dCE was used for MTBDR*plus* and MTBDR*sl* each. MTBDR*plus* and MTBDR*sl* results were reported as described[24]: either actionable [TUB-band positive and determinate (gene-specific locus bands present)] or non-actionable [TUB-band negative or TUB-band positive but indeterminate (gene-specific locus band absent)]. Susceptibility calls were made for all actionable results. Banding patterns were read by two experienced independent readers blinded to each other's calls, the Ultra and Xpert results, and, for the dilution series experiement, the strain antibiograms (if there was a discrepancy between readers, a third experienced reader reviewed results and did the final classification).

Other chambers. MTBDR*plus* and MTBDR*sl* were done on C2 and C4 CEs from both Ultra and Xpert done on the dilution series. C1, C3, and C5 were not tested with LPAs as their CEs were 16S rRNA qPCR-negative or there was no volume remaining to test after the Ultra or Xpert test had completed (C5).

FluoroType MTBDR on cartridge extract. *Diamond chamber.* dCEs from Ultra and Xpert cartridges done on the *in vitro* dilution series were tested by FT using the manufacturer's instructions[25]. A total of 26 tubes for each test (Ultra, Xpert) were tested [four dilutions from 10^1–10^4 CFU/ml in triplicate for both strains plus a negative control for each strain, $(4 \times 3 \times 2 + 2)$]. As Xpert dCE had a volume of ~15 µl, after MTBDR*plus* (5 µl), MTBDR*sl* (5 µl), and the 16S rRNA qPCR (3 µl) were all done on the same Xpert dCE, the remaining volumes (5–14 µl) were made up to 20 µl with dH_2O for FT (the recommended input volume)[25]. All Ultra dCEs (~35 µl originally) had 20 µl remaining and the full 20 µl was used for FT. FT results were classified in a manner similar to that for the line probe assays: actionable (MTBC detected; rifampicin and isoniazid susceptible or resistant) or non-actionable (no MTBC detected, MTBC indeterminate or MTBC detected but rifampicin or isoniazid indeterminate).

Other chambers. FT was done on C2 and C4 (as for LPAs) from both Ultra and Xpert cartridges used for the dilution series.

Evaluation of *rpoB* amplicon cross-contamination risk. *Amplicon escape during batched cartridge extractions.* During all Ultra and Xpert diamond chamber extractions, 1.5 ml microcentrifuge tubes containing 100 µl sterile dH_2O were positioned in the same BSL2 cabinet (Fig. 3A). Three tubes remained open throughout all extractions for each batch extraction and three remained closed (negative controls). Tubes were stored at −20 °C for later FT testing. A total of 20 batches of cartridges were extracted [n = 120 tubes in total from the 20 batches, n = 60 open tubes and n = 60 closed tubes including triplicates], with a median (IQR) number of cartridges per batch of 17.5 (10.5–27.5). There were also three tubes open for each individual cartridge extraction but these were not tested further based on results of the open tubes during batched extraction, which revealed no cross-contamination. Furthermore, extractions procedures were done by a total of five different users to reflect user variability.

Spiking of amplicons. The same XDR-TB strain with known Xpert and Ultra *rpoB* resistance profiles was used in the dilution series (Fig. 1C). Ultra and Xpert were each done on 1 ml of a 10^4 CFU/ml concentration (in triplicate). dCEs were extracted and used for a dilution series (10^0, 10^{-3}, 10^{-6}, and 10^{-9}; each 1 ml final volume). For all dilutions, 5 µl was added to 700 µl of the DS-TB strain (10^4 CFU/ml) and tested with Ultra [700 µl was used as, when combined with the recommended two-fold sample reagent volume, the 2 ml input volume is reached with minimal sample unused (~100 µl)].

Results

***Mycobacterium tuberculosis* complex genomic DNA detection in different chambers from cartridges done on dilution series.** Though qPCR-positive results were obtained from C2, C4 and the dCE (Fig. S1), these results were highly variable even at high concentrations of bacilli (at least 10^4 CFU/ml), suggesting interference. As C2, C4 and dCE gave positive qPCR results on cartridges done on some dilutions, and C1 and C3 gave none, we only explored the utility of the former for downstream testing using MTBDR*plus*, MTBDR*sl*, and FT.

MTBDR*plus* and MTBDR*sl* on extract from cartridges done on dilution series. *TB detection.* More Ultras were MTBC-positive at lower CFU titres than Xpert [e.g., 4/6 (67%) of the 10^1 CFU/ml aliquots vs. 1/6 (17%) for Xpert at the same concentration for both strains] (Fig. 4). MTBDR*plus* had high rates of non-actionable results across all dilutions irrespective of the cartridge chamber extract originated from (diamond, C2, C4) or initial test (Ultra, Xpert) (Fig. 4). MTBDR*sl* had actionable results for all Ultra dCEs $\geq 10^3$ CFU/ml and, for Xpert, all but one dCE $\geq 10^3$ CFU/ml (one Xpert replicate at 10^3 CFU/ml was MTBDR*sl*-non-actionable). MTBDR*sl* on C2 and C4 had non-actionable results across all dilutions (Ultra and Xpert).

Resistance detection. MTBDR*sl* correctly identified FQ and SLID resistance on Ultra dCE done on all XDR strain aliquots $\geq 10^3$ CFU/ml (Fig. 5). On the DS-TB strain, MTBDR*sl* identified FQ susceptibility in all three 10^4 CFU/ml replicates and in 2/3 (67%) replicates for SLIDs (one indeterminate). At 10^3 CFU/ml for the DS-TB strain, 2/3 (67%) were correctly identified as FQ susceptible (one indeterminate) and all were SLID-indeterminate. The C_{Tmin} threshold at which all MTBDR*sl* results was feasible on Ultra CE was <25, which was used for further experiments. Similar results were obtained for MTBDR*sl* on Xpert dCE.

MTBDR*sl* on extract from cartridges done on clinical specimens. *TB detection.* As MTBDR*plus* was not feasible in the *in vitro* assessment, it was not done on CE from Ultras done on clinical sputa. MTBDR*sl*

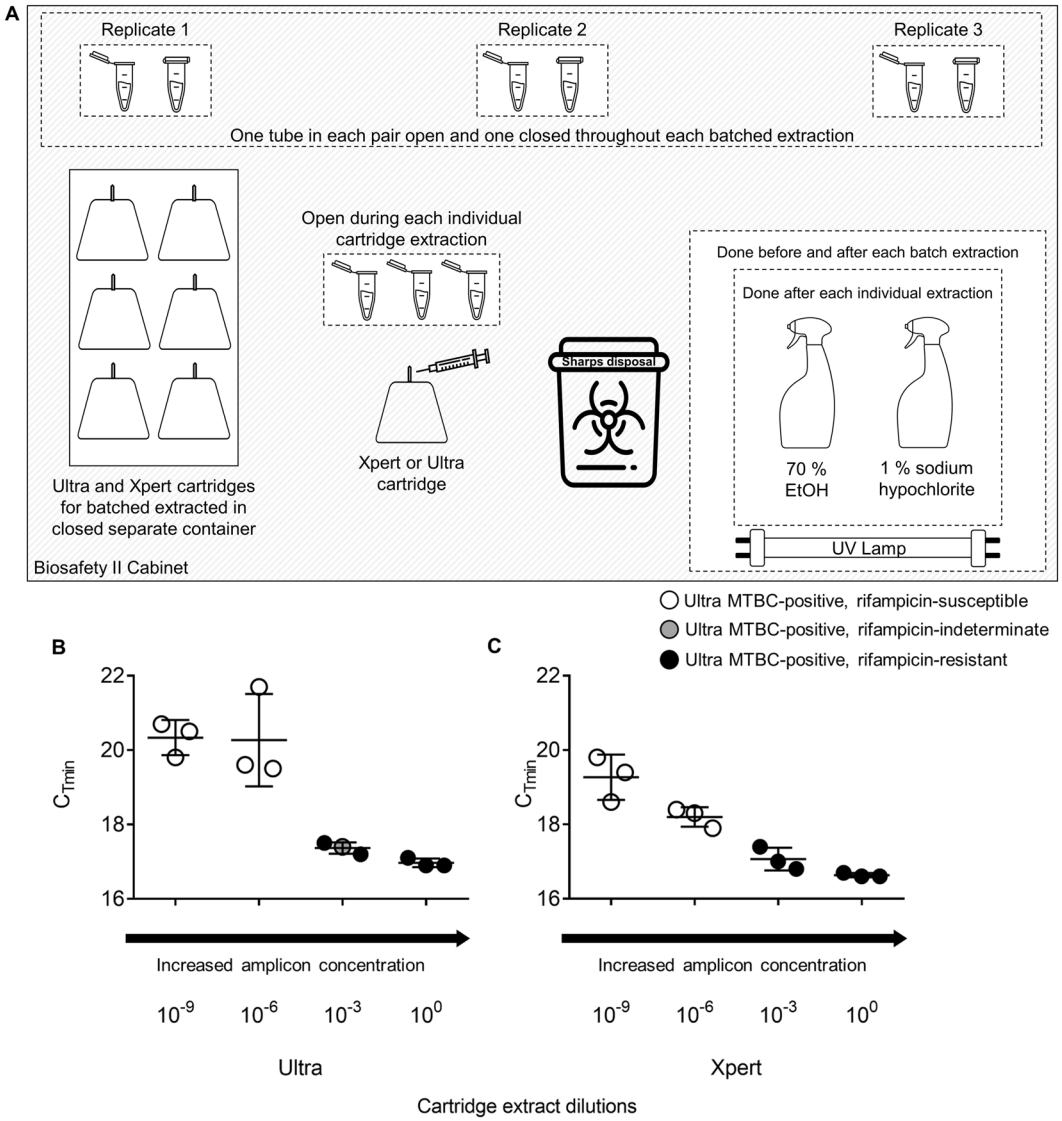


Figure 3. Evaluation of *rpoB* cross contamination risk experimental set-up and results. (**A**) Configuration of the environmental exposure experiment within a Biosafety level 2 cabinet. Three microcentrifuge tubes were open throughout each batched extraction procedure and three remained closed [median (IQR) extractions per batch 17.5 (10.5–27.5)]. No exposed tubes were FT *rpoB*-positive. In parallel to evaluate if, in an absolute worst case scenario, *rpoB* cross-contamination was probable, dCE from a (**B**) Ultra or (**C**) Xpert done on a drug-resistant strain was added to a drug-susceptible strain and the resultant mixture tested by Ultra. When samples of DS-TB contained CE at higher concentrations (undiluted, 10^{-3}), false-resistant (solid black circles) or indeterminate rifampicin resistance (grey circles) are seen. All samples containing CE dilutions beyond 10^{-6} showed true rifampicin susceptibility (white circles). Error bars represent C_{Tmin} values for each dilution. Some images were obtained from the Noun Project: microcentrifuge tube (without changes), Anthony Ledoux, https://thenounproject.com/term/eppendorf/1699532/; spray bottle (without changes), John Winowiecki, https://thenounproject.com/search/?q=spray%20bottle&i=2236898; sharps container, Juicy Fish (with changes), https://thenounproject.com/term/hospital-waste-bin/2450390/; needle (without changes), Creative Mania; https://thenounproject.com/search/?q=injection&creator=2251916&i=2409865.

on dCE from Ultra done on clinical sputa had 37/40 (93%) actionable results (the rest were non-actionable). Non-actionable results corresponded to "trace" or "very low" semi-quantitative categories.

Resistance detection. Of the actionable results, 35/37 (95%) fell within the defined threshold (C_{Tmin} <25) and of these all FQ results were concordant with MTBDR*sl* on sputum and all but one SLID result were concordant (false-susceptible). Though this percentage is slightly higher than the number of patients with C_{Tmin} <25 in our setting, which was determined to be 86% (based on an evaluation of Ultra done in sympotmatic patients in primary care[26]), which further show that this approach would benefit the majority of patients in our setting. Of the 2/37 (5%) results that were actionable but fell above the defined threshold, one was concordant with MTBDR*sl* on sputa and one was indeterminate for FQs and discordant for SLIDs (false-resistant).

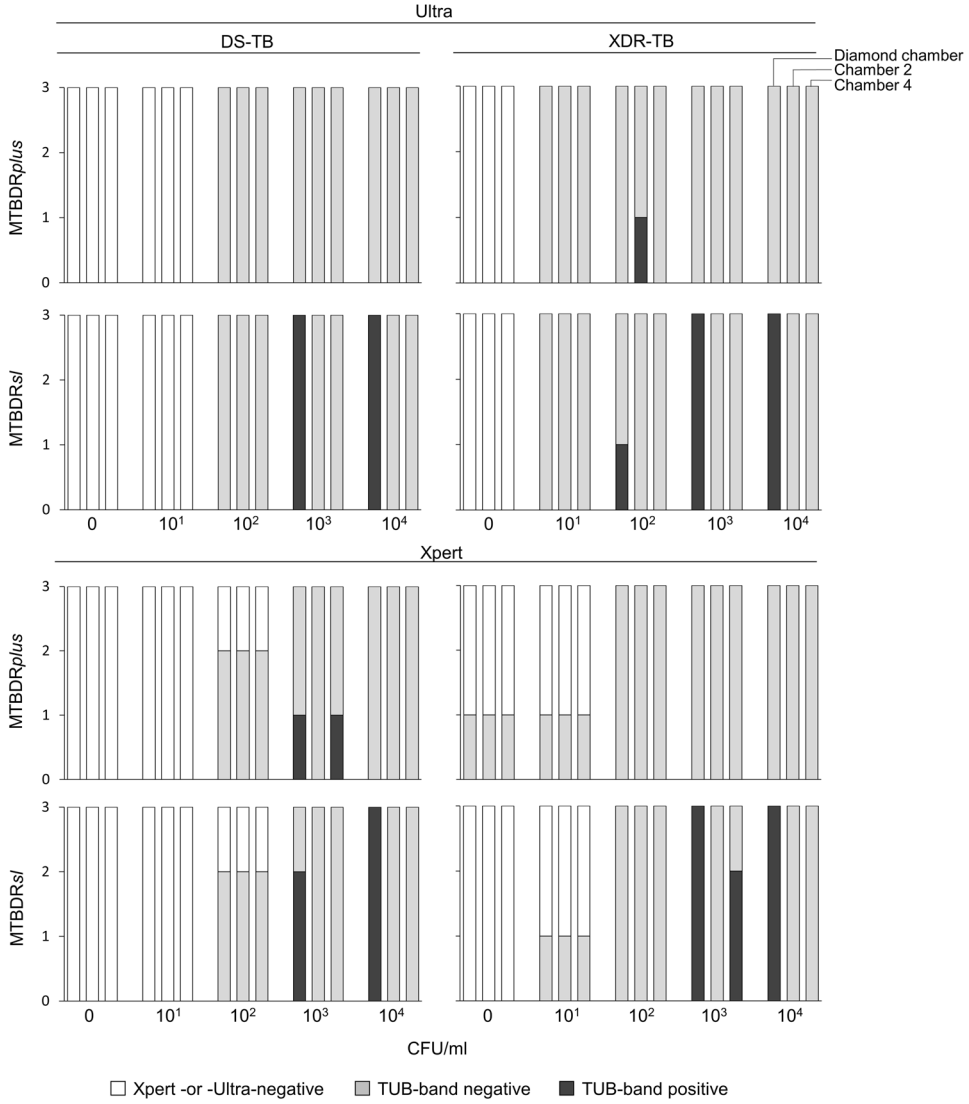


Figure 4. MTBDR*plus* and MTBDR*sl* on cartridge extract results for TB detection. dCE (left-most column), C2 (middle column) and C4 (right-most column) from *M. tuberculosis*-positive cartridges on dilution series (DS-TB and XDR-TB strains) are shown. MTBDR*plus* had mostly non-actionable results (not positive or negative). MTBDR*sl* had actionable results on all Ultra- and Xpert-positive dCE at $>10^3$. Though some actionable line probe assay results for non-diamond chambers were observed, these were inconsistent and had low reproducibility.

Receiver operator curve for determining actionable results. An Ultra *rpoB* C_{Tmin} threshold of <25.4 was defined for dCEs done on clinical sputa with sensitivities of 97% (95% CI 87–100) and specificities of 100% (55–100) (Fig. 6).

FluoroType MTBDR on extract from cartridges done on dilution series. *Diamond chamber.* FT had similar results to MTBDR*plus* on CEs. For example, 3/24 (12%) Ultra dCEs were MTBC-positive (the others negative) for both strains (Fig. S2). In the three Ultra dCEs with a TB-positive FT result, all had indeterminate susceptibility results for at least one drug. A total of 18/24 (75%) Xpert dCEs were FT MTBC-positive, however, of these 13/24 (54) were indeterminate for at least one drug.

Chamber 2. FT on Ultra C2 had MTBC positivity rates of 10/12 (83%) and 11/12 (92%) for DS-TB and XDR-TB, respectively. On Xpert C2, FT TB positivity rates were 5/12 (42%) and 7/12 (58%) for DS-TB and XDR-TB, respectively. In MTBC-positive extracts (Ultra and Xpert), most resistance calls were indeterminate or discordant with the paired isolate.

Chamber 4. FT done on C4 from Ultra had 8/12 (67%) and 9/12 (75%) TB positivity rates for DS-TB and XDR-TB strains respectively, and 3/12 (25%) and 1/12 (8%) on for C4 from Xpert respectively. As for C2, resistance calls were mainly indeterminate or discordant with paired isolate.

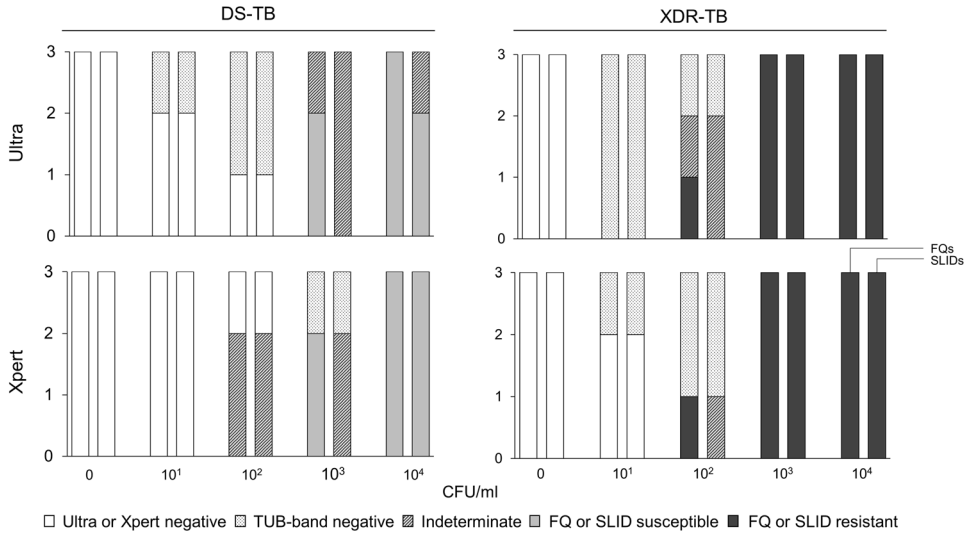


Figure 5. MTBDR*sl* drug susceptibility results on dCEs from Ultra and Xpert on dilution series. All results $\geq 10^3$ CFU/ml for the XDR-TB strain had resistance results concordant with the isolate. Some SLIDs indeterminate results were seen for the DS-TB $>10^3$ at the same concentrations but MTBDR*sl* results were otherwise concordant with those on the isolate.

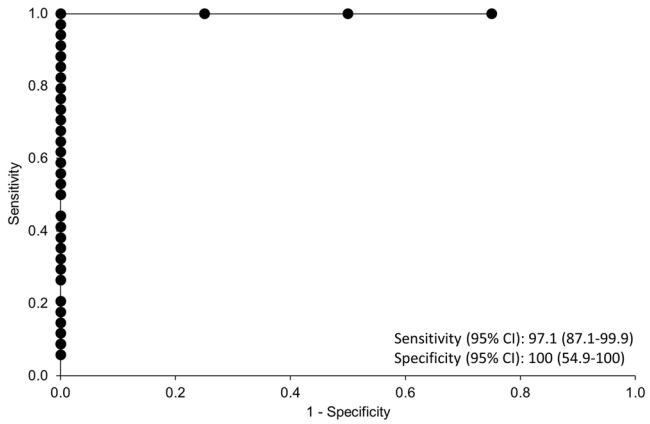


Figure 6. Receiver operation area under the curve of actionable vs. non-actionable results of MTBDR*sl* on Ultra diamond cartridge extract done on DR-TB clinical sputa to determine a C_{Tmin} threshold at which this approach is not feasible. MTBDR*sl* yields actionable results on cartridge extract from Ultra at a C_{Tmin} threshold of <25.4 with a sensitivity of 97% (87.1–99.9; 95% CI) and specificity of 100% (54.9–100; 95% CI) respectively.

***rpoB* amplicon cross-contamination risk evaluation.** *Exposure of open tubes during batched extractions.* All sixty tubes exposed were FT MTBC-negative and had no *rpoB* amplification.

Amplicon spiking for absolute worst-case cross-contamination scenario. Of the Ultra dCEs done on XDR-TB and spiked into DS-TB for re-testing with Ultra, evidence of cross-contamination was seen when dCEs were diluted less than 10^{-6} before addition to the DS-TB strain [3/3 (100%) of 10^0 dilutions and 2/3 (67%) of the 10^{-3} dilutions showed false-resistance (1/3 of the 10^{-3} was resistance indeterminate)] (Fig. 3B). Similar results were obtained for Xpert dCE (Fig. 3C).

Discussion

We have validated MTBDR*sl* on CEs from used Ultra cartridges for genotypic second-line DST. We show: (1) MTBDR*sl* on Ultra dCE when C_{Tmin} <25 enabled DST concordant with sputum results, (2) risk of *rpoB* extract cross-contamination is unlikely if standard aseptic protocols are followed, (3) neither 16S rRNA qPCR, MTBDR*plus*, MTBDR*sl* nor FT are feasible on other cartridge chambers, nor was MTBDR*plus* or FT on Ultra and Xpert dCEs. These data support the use of Ultra extract for second-line genotypic DST.

We defined a threshold at which MTBDR*sl* is likely to work on Ultra dCE from the vast majority of Ultra-positive patients, thereby avoiding time and resources wasted on dCE unlikely to give a valid result. We are mindful that there were some indeterminate SLID results (in line with previous reports of higher MTBDR*sl* indeterminate result rates for SLIDs vs. FQs)[27–29]. However, all dCE SLID-indeterminate results from the dilution series

were from the DS-TB strain and there were no indeterminate SLID results on XDR-TB dCEs. On clinical sputum (and falling within our threshold), one MTBDR*sl* SLID susceptibility result was discordant with sputum (one false-negative). We thus suggest that MTBDR*sl* Ultra dCE results are interpreted in the same manner as recommended by the WHO for MTBDR*sl* on clinical specimens[30]. If, for example, MTBDR*sl* on dCE is non-actionable or susceptible, MTBDR*sl* on sputum or isolates should be done. If there is still no evidence of resistance in a high burden setting, phenotypic DST should still be done given the suboptimal rule-out accuracy of MTBDR*sl*[19,30].

The possibility of contamination from *rpoB* amplicons during extractions has not been investigated. We implemented systematic testing for possible environmental contamination. No tubes exposed for each extraction batch were *rpoB*-positive when tested with FT. FT was used for testing for *rpoB* amplicons as it is more sensitive than MTBDR*plus*[14,15].

We further tested a worst-case contamination scenario with dCEs from both Ultra and Xpert cartridges done on a XDR-TB strain, diluting these dCEs, and adding them to a DS-TB strain which was subsequently tested by Ultra. The undiluted and most concentrated dCE dilutions (10^0, 10^{-3}) showed false rifampicin-resistance indicating that, although the GeneXpert platform does have proven ability to remove large numbers of amplicons[31], it was not able to remove all amplicons during the pre-amplification wash steps, however, amplicons diluted beyond 10^{-3} were successfully removed to the point of not being detected[22,32,33]. These results, together with those from the environmental samplings during extractions, shows that when standard aseptic techniques are used, amplicon cross-contamination is highly unlikely except in the artificial worst case scenarios. Finally, it should be noted that, in line with good practice in any molecular biology laboratory providing results for patient management, dCEs should not be collected in the same room where *rpoB*-based tests are done, and that the risk of cross-contamination from the dCE approach is only pertinent to tests for rifampicin resistance.

We suggest that diagnosticians considering implementing this approach use the cartridge itself as a transport vessel (upright and in sealed containers) to a central laboratory where dCE can be extracted appropriately (the diamond is a sealed chamber and should remain safe during transport). Most peripheral laboratories will be unable to do the dCE procedure safely and downstream molecular DST like MTBDR*sl*. This cartridge transport can interface with existing specimen referral networks. If dCE is planned purely for molecular epidemiology, we suggest that dCE be extracted and stored at −80 °C or alternatively the whole cartridge be stored at −20 °C until extractions can be done in a batched, centralised fashion. The long term stability of these approaches will require examination.

We further hypothesised that liquid from other cartridge chambers may avoid interference by *rpoB* amplicons. However, upon testing, this approach gave variable non-replicable results. This was true for qPCR, MTBDR*plus*, MTBDR*sl* and FT assays. This may also be due to very low concentrations of template in these chambers, for example C3 – which is the "wash chamber", and/or remnant PCR inhibitors (e.g., salts from the sample reagent). In light of this, we believe that the presence of these amplicons may prevent newer approaches, such as next generation sequencing methods, from performing well on dCE without to clean up steps. This warrants further investigation. CE from the diamond chamber hence remains the best option for downstream genotypic DST.

The results of this study should be interpreted within its limitations, namely aseptic techniques done in an assay- or procedure-specific biosafety cabinet are needed to minimise amplicon cross-contamination. However, this infrastructure should already be implemented per WHO guidelines[34] where LPAs are done routinely for patient care. Furthermore, per good laboratory practice, CEs should not be collected in the same room where *rpoB*- or *IS6110/1081*-based assays are done, nor should either procedure be done by the same personnel on a daily basis. Lastly, further investigation into cross-contamination risk should be done in a routine diagnostic setting. This should include multiple operators.

We also acknowledge that this method may increase risk of needle stick injury. Standard biosafety protocols should be strictly adhered to. We were recently funded to develop a device that can eject material from cartridges in a safe manner. Another limitation is MTBDR*plus* was not feasible on Ultra CEs and we suspect this is due to interference from both *rpoB* and IS*6110*/*1081* amplicons. Thus, combined with the large volumes (and hence diluted targeted DNA) recovered from non-diamond chambers in Ultra and Xpert, MTBDR*plus* (and also likely FT) on extract from any Ultra cartridge chamber is in all likelihood not useful for isoniazid or confirmatory rifampicin DST. Finally, although the diamond chamber is a closed system and appears protected against desiccation, we acknowledge that some desiccation may occur over prolonged periods that this requires future systematic evaluation. However, we recommend that extract method is done on an as fresh a cartridge as possible (either at a peripheral or central laboratory), in order to reduce the delays of DR-TB diagnosis. Formal evaluation of CE stability pre-extraction may be useful.

We conclude that dCEs from Ultra at the C_{Tmin} threshold (<25), can be used for genotypic second-line DST (MTBDR*sl*). Ultra and MTBDR*sl* on dCE therefore allows for the rapid rule-in detection of XDR-TB on a single specimen.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on request.

Received: 8 September 2019; Accepted: 24 January 2020;
Published online: 14 February 2020

References

1. Dheda, K. *et al.* The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. *The lancet Respiratory medicine* **5**, 291–360 (2017).
2. World Health Organization. *Global Tuberculosis Report* (2019).
3. Gandhi, N. R. *et al.* Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *The Lancet* **375**, 1830–1843 (2010).

4. Naidoo, P. *et al*. The South African Tuberculosis Care Cascade: Estimated Losses and Methodological Challenges. *The Journal of infectious diseases* **216**, S702–S713 (2017).
5. Subbaraman, R. *et al*. The tuberculosis cascade of care in India's public sector: a systematic review and meta-analysis. *PLoS medicine* **13**, e1002149 (2016).
6. Venter, R. *et al*. Mycobacterial genomic DNA from used Xpert MTB/RIF cartridges can be utilised for accurate second-line genotypic drug susceptibility testing and spoligotyping. *Scientific reports* **7**, 14854 (2017).
7. Theron, G. *et al*. The Use of an Automated Quantitative Polymerase Chain Reaction (Xpert MTB/RIF) to Predict the Sputum Smear Status of Tuberculosis Patients. *Clinical Infectious Diseases* **54**, 384–388 (2012).
8. Alame-Emane, A. K. *et al*. The use of GeneXpert remnants for drug resistance profiling and molecular epidemiology of tuberculosis in Libreville, Gabon. *Journal of Clinical Microbiology, JCM*. 02257–02216 (2017).
9. Mambuque, E. T. *et al*. Direct genotyping of Mycobacterium tuberculosis from Xpert® MTB/RIF remnants. *Tuberculosis* (2018).
10. World Health Organization. WHO meeting report of a technical expert consultation: Non-inferiority analysis of Xpert MTB/RIF Ultra compared to Xpert MTB/RIF. (2017).
11. Chakravorty, S. *et al*. The new Xpert MTB/RIF Ultra: improving detection of Mycobacterium tuberculosis and resistance to rifampin in an assay suitable for point-of-care testing. *MBio* **8**, e00812–00817 (2017).
12. Dorman, S. E. *et al*. Xpert MTB/RIF Ultra for detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. *The Lancet Infectious Diseases* **18**, 76–84 (2018).
13. Mishra, H. *et al*. Diagnostic accuracy and predictive value of Xpert Ultra and Xpert MTB/RIF for tuberculosis diagnosis in an HIV-endemic setting with a high burden of previous tuberculosis Lancet Respiratory Medicine (2019).
14. de Vos, M. *et al*. Diagnostic accuracy and utility of FluoroType MTBDR, a new molecular assay for multidrug-resistant tuberculosis. *Journal of clinical microbiology* **56**, e00531–00518 (2018).
15. Hillemann, D., Haasis, C., Andres, S., Behn, T. & Kranzer, K. Validation of the FluoroType MTBDR assay for detection of rifampin and isoniazid resistance in Mycobacterium tuberculosis complex isolates. *Journal of clinical microbiology* **56**, e00072–00018 (2018).
16. World Health Organization. Xpert MTB/RIF implementation manual: technical and perational 'how-to': practical considerations [Internet]. Geneva: World Health Organization. (2014).
17. Organization, W. H. WHO meeting report of a technical expert consultation: non-inferiority analysis of Xpert MTB/RIF Ultra compared to Xpert MTB/RIF. (World Health Organization, 2017).
18. Hain Lifescience. GenoType MTBDRsl VER 2.0 Instructions for Use. (2015).
19. Motsoaledi, A. *Health Department of the Republic of South Africa* (2014).
20. Cepheid. Journey Inside the Cepheid GeneXpert® Cartridge - 3D Animation (2019).
21. Raja, S. *et al*. *Technology for automated, rapid, and quantitative PCR or reverse transcription-PCR clinical testing*. **51**, 882–890 (2005).
22. Theron, G. In *Molecular Microbiology: Diagnostic Principles and Practice* (ed. David H. Persing *et al*.) Ch. 40, (ASM Press, 2016).
23. Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of microbiological methods* **69**, 330–339 (2007).
24. Derendinger, B. *et al*. *Widespread use of incorrect PCR ramp rate negatively impacts multidrug-resistant tuberculosis diagnosis (MTBDR plus)*. **8**, 3206 (2018).
25. Hain Lifescience. FluoroType® MTBDR VER 2.0 Instructions for Use. (2019).
26. Mishra, H. *et al*. Xpert MTB/RIF Ultra and Xpert MTB/RIF for diagnosis of tuberculosis in an HIV-endemic setting with a high burden of previous tuberculosis: a two-cohort diagnostic accuracy study. *Lancet Respiratory Medicine*, https://doi.org/10.1016/S2213-2600(19)30370-4 (2020).
27. Mao, X. *et al*. Diagnosis of drug resistance to fluoroquinolones, amikacin, capreomycin, kanamycin and ethambutol with Genotype MTBDRsl assay: a meta-analysis. *Annals of Clinical & Laboratory Science* **45**, 533–544 (2015).
28. Theron, G. *et al*. The diagnostic accuracy of the GenoType (®) MTBDRsl assay for the detection of resistance to second-line anti-tuberculosis drugs. *Cochrane Database of Systematic Reviews* **10** (2014).
29. Tomasicchio, M. *et al*. The diagnostic accuracy of the MTBDRplus and MTBDRsl assays for drug-resistant TB detection when performed on sputum and culture isolates. *Scientific reports* **6** (2016).
30. World Health Organization. WHO treatment guidelines for drug-resistant tuberculosis. (2016).
31. Blakemore, R. *et al*. Evaluation of the analytical performance of the Xpert MTB/RIF assay. *Journal of clinical microbiology* **48**, 2495–2501 (2010).
32. Theron, G. *et al*. Xpert MTB/RIF results in patients with previous tuberculosis: can we distinguish true from false positive results? *Clinical Infectious Diseases* **62**, 995–1001 (2016).
33. Theron, G. *et al*. False positive Xpert MTB/RIF results in re-tested patients with previous tuberculosis: frequency, profile, and prospective clinical outcomes. *Journal of clinical microbiology*, JCM. 01696–01617 (2018).
34. World Health Organization. The use of molecular line probe assay for the detection of resistance to isoniazid and rifampicin: policy update. (2016).

Acknowledgements

The authors would like to sincerely thank Roxanne Higgit for her assistance in the extraction process as well as Dr. Charissa Naidoo for assistance with the qPCR protocol and Dr Anzaan Dippenaar for her assistance with the FT assays. The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author(s) and are not necessarily to be attributed to the NRF. Research reported in this publication was supported by the South African Medical Research Council. The content is solely the responsibility of the authors and does not necessarily represent the official views of the South African Medical Research Council. GT acknowledges funding from South African Medical Research Council (SAMRC Flagship Project MRC-RFA-IFSP-01-2013), the EDCTP2 program supported by the European Union (grant SF1401, OPTIMAL DIAGNOSIS), and the Faculty of Medicine and Health Sciences, Stellenbosch University. This research was supported by The Center for Innovation in Point-of-Care Technologies for HIV/AIDS at NorthwesternUniversity. The Research reported in this publication was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award Number U54EB027049. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

G.T., R.W., M.D.V. and R.V. conceived the experiments. R.V., S.M., B.D., H.T., and A.R. conducted the experiments. T.D. provided specimens and data from the NHLS. R.V. and S.M. analysed data. All authors reviewed the manuscript.

Competing interests

G.T. acknowledges in-kind donations from Hain Lifesciences. G.T. acknowledges in-kind donations from Cepheid for other studies. G.T., R.W., and M.d.V. acknowledge in-kind donations and funding from Hain Lifesciences for other studies. G.T., R.W. and M.d.V. declare no other competing interests. The other authors each declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-59164-3.

Correspondence and requests for materials should be addressed to G.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix VI

Frequent suboptimal thermocycler ramp rate usage negatively impacts GenoType MTBDR*sl* VER 2.0 performance for second-line drug resistant tuberculosis diagnosis

Publication status: published

Derendinger, B., de Vos, M., Pillay, S., Venter, R., Metcalfe, J., Ghebrekristos, Y., **Minnies, S.**, Dolby, T., Beylis, N., Warren, R. and Theron, G., 2021. Frequent suboptimal thermocycler ramp rate usage negatively impacts MTBDRsl performance for second-line drug resistant tuberculosis diagnosis. *The Journal of Molecular Diagnostics*, *24*(5), pp.494-502.

DOI: https://doi.org/10.1016/j.jmoldx.2022.01.003

The Journal of Molecular Diagnostics, Vol. 24, No. 5, May 2022

the Journal of Molecular Diagnostics
jmdjournal.org

Frequent Suboptimal Thermocycler Ramp Rate Usage Negatively Impacts GenoType MTBDR*sl* VER 2.0 Performance for Second-Line Drug-Resistant Tuberculosis Diagnosis

Brigitta Derendinger,* Margaretha de Vos,* Samantha Pillay,* Rouxjeane Venter,* John Metcalfe,† Yonas Ghebrekristos,*‡ Stephanie Minnies,* Tania Dolby,§ Natalie Beylis,§ Robin Warren,* and Grant Theron*

From the DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, SA-MRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa; the Division of Pulmonary and Critical Care Medicine,† Zuckerberg San Francisco General Hospital, University of California San Francisco, San Francisco, California; the National Health Laboratory Service,‡ Groote Schuur Hospital, Cape Town, South Africa; and the National Health Laboratory Service,§ Green Point, Cape Town, South Africa*

Accepted for publication January 13, 2022.

Address correspondence to Grant Theron, Ph.D., Faculty of Health Sciences, Department of Molecular and Biology and Human Genetics, BMRI Bldg., 2nd Floor, Room 2035, Francie can Zijl Dr., Tygerberg Campus, Tygerberg 7505, South Africa. E-mail: gtheron@sun.ac.za.

Strengthening second-line drug-resistant tuberculosis (TB) detection is a priority. GenoType MTBDR*plus* VER 2.0 performance is reduced with non-recommended ramp rate usage (temperature change speed between PCR cycles); however, ramp rate's effect on GenoType MTBDR*sl* VER 2.0 (MTBDR*sl*) performance, is unknown. Fifty-two Xpert MTB/RIF Ultra-positive rifampicin-resistant smear-negative sputa and a *Mycobacterium tuberculosis* dilution series were tested at a manufacturer-recommended (2.2°C/second) or suboptimal (4.0°C/second) ramp rate. *M. tuberculosis*–complex-DNA positivity, indeterminates, fluoroquinolone- and second-line injectable-resistance accuracy, banding differences, and, separately, inter-reader variability were assessed. Five (39%) of 13 re-surveyed laboratories did not use the manufacturer-recommended ramp rate. On sputum, 2.2°C/second improved indeterminates versus 4.0°C/second (0 of 52 versus 7 of 51; $P = 0.006$), incorrect drug-class diagnostic calls (0 of 104 versus 6 of 102; $P = 0.013$), and incorrect banding calls (0 of 1300 versus 54 of 1275; $P < 0.001$). Similarly, 2.2°C/second improved valid results [(52 of 52 versus 41 of 51; +21% ($P = 0.001$)] and banding call inter-reader variability [34 of 1300 (3%) versus 52 of 1300 (4%); $P = 0.030$]. At the suboptimal ramp rate, false-resistance and false-susceptible calls resulted from wild-type band absence rather than mutant band appearance, resulting in misclassification of moxifloxacin resistance level from high-to-low. Suboptimal ramp rate contributes to poor MTBDR*sl* performance. Laboratories must ensure that the manufacturer-recommended ramp rate is used. *(J Mol Diagn 2022, 24: 494–502; https://doi.org/10.1016/j.jmoldx.2022.01.003)*

Supported by the European and Developing Countries Clinical Trials (EDCTP2) program supported by the European Union (grant SF1401, OPTIMAL DIAGNOSIS), The Center for Innovation in Point-of-Care Technologies for HIV/AIDS at Northwestern University (National Institute of Biomedical Imaging and Bioengineering of the NIH, Award Number U54EB027049), and the National Institute of Allergy and Infectious Diseases. Research reported in this publication was also supported by the South African Medical Research Council.

Disclosures: Hain Lifescience donated the MTBDR*sl* kits used for the dilution series. They did not have any role in study design or result interpretation.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the South African Medical Research Council.

Current address of M.d.V., Foundation for Innovative New Diagnostics, Geneva, Switzerland.

https://doi.org/10.1016/j.jmoldx.2022.01.003

In 2019, approximately 10 million individuals fell ill with tuberculosis (TB) and approximately 1.3 million individuals died.[1] Drug-resistant TB is a global health problem. Approximately 465,000 individuals having multidrug-resistant TB (MDR-TB), ≥6% of whom have additional resistance to fluoroquinolones (FQs) and second-line injectables (SLIDs) (WHO Global Tuberculosis Report 2020). Worldwide in 2019, only 52% of patients with MDR-TB were tested for resistance to both these drug classes, and only 58% of those who start treatment successfully complete it (WHO Global Tuberculosis Report 2020). Phenotypic culture-based drug susceptibility testing is slow and costly, and patients need to wait up to 6 months before being placed on effective treatment, if at all.[2] FQs are becoming incorporated into first-line drug regimens, which will require drastic scale-up of drug susceptibility testing. The World Health Organization (WHO) also recommends moxifloxacin for isoniazid-monoresistant TB in the newly endorsed shortened rifapentine regimen.[3]

GenoType MTBDR*sl* VER 2.0 (MTBDR*sl*) (Hain Lifescience, Nehren, Germany) is one of two commercially available rapid molecular WHO-endorsed assays for the detection of *Mycobacterium tuberculosis* complex and resistance to FQs and SLIDs.[4,5] According to the WHO, MTBDR*sl* should be performed directly on sputum irrespective of smear microscopy status to reduce the delay associated with culture for indirect testing.[4]

However, performance data for direct use on sputum are heterogeneous. In a systematic review and meta-analysis, smear-negative sensitivity estimates were imprecise: 80% [95% CI, 28−99], 80% (95% CI, 28−99), and 50% (95% CI, 1−99) for FQs, SLIDs, and extensively drug-resistant TB (XDR-TB) (using the then contemporaneous definition), respectively.[6] This affected the certainty of evidence of the WHO recommendation and undermined uptake of MTBDR*sl*.

MTBDR*sl* requires thermocycling for DNA amplification. The manufacturer recommends a ramp rate of ≤2.2°C/second, which is the speed of temperature change between PCR cycles. It was previously shown that performance of GenoType MTBDR*plus* VER 2.0 (MTBDR*plus*) (Hain Lifescience), which is an assay for first-line resistance, is reduced when suboptimal thermocycler ramp rates are used, mainly on smear-negative specimens.[7] These findings are incorporated into laboratory external quality assessment programs and the WHO TB laboratory training material (*https://openwho.org/courses/multi-drug-resistant-tb*, last accessed July 6, 2021).

If MTBDR*sl* is also vulnerable to this phenomenon, this would result in some of the thousands of individuals who receive this assay each day having drug resistance diagnoses missed, thereby resulting in resistance to the drugs critical to protect new regimens (eg, FQ to limit bedaquiline resistance acquisition in the oral second-line regimen) remaining delayed or undiagnosed.[8,9] More broadly, this issue of ramp rate is increasingly pertinent as manufacturers are designing instruments with faster thermocycling (and hence faster ramp rates) to decrease time-to-result. Furthermore, many thermocyclers, especially those at entry level (ie, with fewer customizable settings compared with more advanced models that are typically more expensive), do not have a customizable ramp rate.

It is hypothesized that the heterogeneous and suboptimal sensitivities reported for MTBDR*sl* on smear-negative specimens were partly attributable to suboptimal ramp rate, and the goal was to generate empirical evidence of this theory. The current study assessed whether laboratories that reported use of suboptimal ramp rates during the authors' previous MTBDR*plus* evaluation[7] had switched to the manufacturer-recommended ramp rate and what the observed effect had been.

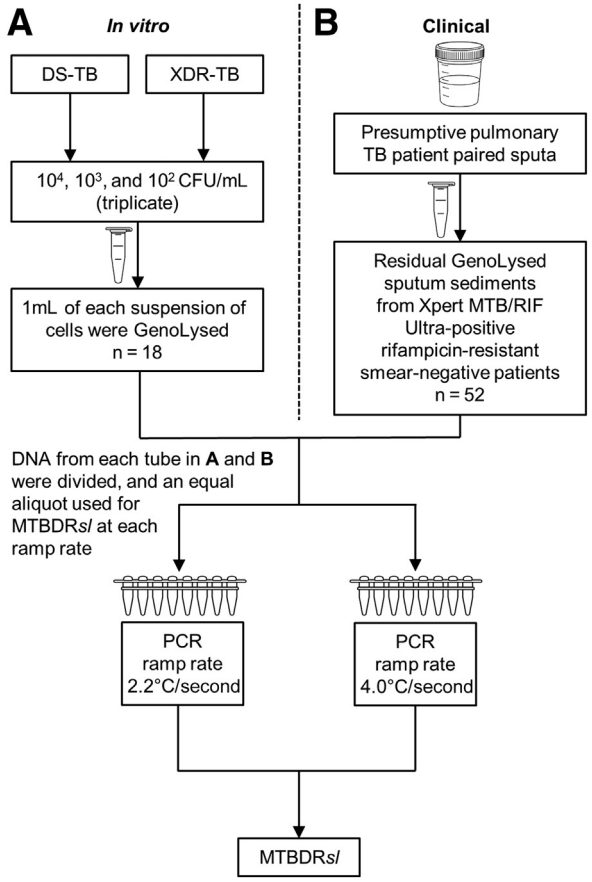


Figure 1 Study flow diagram for an *in vitro* [a dilution series of cells (10^4, 10^3, and 10^2 colony-forming units per milliliter [CFU/mL])] experiment (**A**) and clinical experiment (sputa) (**B**) to assess the impact of thermocycler ramp rate on GenoType MTBDR*sl* VER 2.0 (MTBDR*sl*). DNA extracted from the dilution series and clinical specimens was split and MTBDR*sl* compared head-to-head at the manufacturer-recommended ramp rate of 2.2°C/second or 4.0°C/second. DS-TB, drug-susceptible tuberculosis; TB, tuberculosis; XDR-TB, extensively drug-resistant tuberculosis.

Table 1 MTBDR*sl* Performance on a Dilution Series of Drug-Susceptible-TB and XDR-TB Strains (10^4, 10^3, and 10^2 CFU/mL) at Ramp Rates of 2.2°C/second (Manufacturer-Recommended) or 4.0°C/second (3 Replicates in Triplicate for Each Ramp Rate; 18 Total MTBDR*sl* Results)

		TUB-band—positive			
Ramp rate (°C/second)	TUB-band—positive	Indeterminate for any gene locus	Incorrect banding call	Incorrect drug class diagnostic call	Valid result
2.2	16/18* (89)	2/16† (13)	22/400‡ (6)	2/32§ (6)	14/16† (88)
4.0	17/18* (94), $P = 0.547$	3/17† (18), $P = 0.680$	33/425¶ (8), $P = 0.193$	2/34‖ (6), $P = 0.950$	14/17† (82), $P = 0.680$

Data are expressed as *n/N* (%). Accuracy for *M. tuberculosis*—complex-DNA (TUB-band) and then further analysis of indeterminate rates, incorrect banding calls, and incorrect drug class diagnostic calls were done. No significant differences were seen between ramp rates using dilution series. *P* values are for within-column comparisons between different ramp rates. CFU, colony-forming units; Incorrect banding call, the presence or absence of a band deviating from the true banding call; Incorrect drug class diagnostic call, the presence or absence of banding patterns resulting in deviation of the true susceptibility to a drug class; Indeterminate, one or more gene locus control is absent; MTBDR*sl*, GenoType MTBDR*sl* VER 2.0; TB, tuberculosis; TUB-band—positive, positive for *Mycobacterium tuberculosis*—complex-DNA; Valid result, TUB-band—positive, determinate for all gene locus controls, thus having diagnostic calls for both drug classes; XDR, extensively drug resistant.

*Two strains × 3 replicates × 3 dilutions.

†TUB-positive strips.

‡Sixteen TUB-band—positive strips × 25 bands per strip.

§Sixteen TUB-band—positive strips × 2 drug class diagnostic calls.

¶Seventeen TUB-band—positive strips × 25 bands per strip.

‖Seventeen TUB-band—positive strips × 2 drug class diagnostic calls.

Materials and Methods

Ethics Statement

This study was approved by the Health Research Ethics Committee of Stellenbosch University (N16/04/045) and Western Cape Research Ethics Committee (WC_2016RP18_637). All methods were in accordance with relevant guidelines and regulations. Permission was granted to access anonymized residual specimens collected as part of routine diagnostic practices, and thus patient informed consent was waived.

Experimental Design

Ramp rate assessment was performed in both an *in vitro* dilution series and clinical sputa (Figure 1). DNA extracted from dilution series and clinical specimens were split and compared head-to-head at the manufacturer-recommended ramp rate of 2.2°C/second or the most common suboptimal ramp rate of 4.0°C/second identified previously in a survey.[7] MTBDR*sl* was performed on all amplified DNA per manufacturer's instructions for use (Hain Lifescience) [kit lot #39B (expiry date September 2, 2019); strip lot #ABB0117A161 (expiry date September 18, 2019)]. All experiments for this study were performed before the kits' expiration dates. Strips were interpreted by using the WHO-endorsed Global Laboratory Initiative line probe assay interpretation guide (GLI, *http://www.stoptb.org/wg/gli/assets/documents/LPA_test_web_ready.pdf*; WHO, *https://openwho.org/courses/multi-drug-resistant-tb/items/49CT8rhOFxxXzbJYsIIZlK*, last accessed October 19, 2021) and the authors agree with the recommendations in these guidelines. For sputa, programmatic MTBDR*sl* results (performed at the recommended ramp rate) were compared. All equipment is annually calibrated and serviced.

MTBDR*sl* Calls and Result Definitions

Conjugate Control Band

The conjugate control (CC)-band must be present for a strip to be valid as it indicates that hybridization occurred.

Amplification Control Band

The amplification control (AC)-band is present when the assay is performed correctly. Per the manual (GenoType MTBDR*sl* Instructions for Use IFU-317A-04; Hain Lifescience), there are rare cases in which the AC-band disappears due to competition during the amplification reaction. In this scenario, an absent AC-band in combination with *M. tuberculosis*—complex-DNA (TUB-band) and locus control bands is still a valid result.

Locus Control Bands (*gyrA*, *gyrB*, *rrs*, and *eis*)

The locus control bands (*gyrA*, *gyrB*, *rrs*, and *eis*) need to be present for a call from that locus to not be indeterminate.

Positive for *M. tuberculosis*—Complex-DNA

The TUB-band indicates the presence of *M. tuberculosis*—complex-DNA.

Strip Banding Call

For a band to be classified as present, it must be equal or darker than the AC-band. Overall, there are 27 possible strip bands on MTBDR*sl*. When only the CC- and AC-bands are present, this represents a valid TUB-negative result.

Table 2 MTBDR*sl* Performance on Smear-Negative Sputa at Ramp Rates of 2.2°C/second (Manufacturer-Recommended) or 4.0°C/second (52 Isolates)

		TUB-band—positive			
Ramp rate (°C/second)	TUB-band—positive	Indeterminate for any gene locus	Incorrect banding call	Incorrect drug class diagnostic call	Valid result
2.2	52/52* (100)	0/52† (0)	0/1300‡ (0)	0/104§ (0)	52/52† (100)
4.0	51/52* (98), $P = 0.315$	7/51† (14), **$P = 0.006$**	54/1275¶ (4), **$P < 0.001$**	6/102‖ (6), **$P = 0.013$**	41/51† (80), **$P = 0.001$**

Data are expressed as *n/N* (%). Accuracy for *Mycobacterium tuberculosis*—complex-DNA, and then further analysis of indeterminate rates, incorrect banding calls, and incorrect drug class diagnostic calls were done. The number of valid results [52 of 52 (100%) versus 41 of 51 (80%)] improved by 21% (95% CI, 8—34; $P < 0.001$). *P* values are for within-column comparisons between different ramp rates. Significant *P* values are marked in bold. Incorrect banding call, the presence or absence of a band deviating from the true banding call; Incorrect drug class diagnostic call, the presence or absence of banding patterns resulting in deviation of the true susceptibility to a drug class; Indeterminate, one or more gene locus control is absent; MTBDR*sl*, GenoType MTBDR*sl* VER 2.0; TB, tuberculosis; TUB-band—positive, positive for *Mycobacterium tuberculosis*—complex-DNA; Valid result, TUB-band—positive, determinate for all gene locus controls, thus having diagnostic calls for both drug classes.

*Total number of clinical specimens.

†TUB-positive strips.

‡Fifty-two TUB-band—positive strips × 25 bands per strip.

§Fifty-two TUB-band—positive strips × 2 drug class diagnostic calls.

¶Fifty-one TUB-band—positive strips × 25 bands per strip.

‖Fifty-one TUB-band—positive strips × 2 drug class diagnostic calls.

Drug Class Diagnostic Call

Band presence or absence in a gene region determines whether the result is classified as susceptible or resistance to a drug class (two drug class diagnostic calls possible for MTBDR*sl*: FQs or SLIDs).

(In)determinate for a Gene Region and/or Drug Class

For a specific gene region and/or drug class to be determinate, locus control band(s) must be present. A strip was called indeterminate for a drug class if at least one gene locus control was absent.

Valid Result

TUB-band—positive strip determinate for all gene locus controls and thus has diagnostic calls for both drug classes (eg, TUB-band—positive, FQ-resistant, SLID-susceptible).

Additional Amikacin Resistance (*rrs* C1402T and *eis* C-14T)

These are new guidelines released by the WHO indicating resistance to amikacin. *rrs* C1402T translates to *rrs* WT1 band not binding and *eis* C-14T translates to the *eis* MUT1 band binding.[10] The MTBDR*sl* will need to be updated.

Impact of Thermocycler Ramp Rate on MTBDR*sl* Performance on a Dilution Series

A phenotypically and genotypically resistant clinical XDR strain (*gyrA* D94N, *gyrB* wild type, *rrs* A1401G, and *eis* wild type) and a drug-susceptible strain (H37Rv, ATCC 25618) were grown to mid-exponential phase in Middlebrook 7H9 media (Becton Dickinson, Franklin Lakes, NJ) supplemented with Middlebrook Oleic Albumin Dextrose Catalase (Becton Dickinson) and adjusted to a McFarland 1.0 standard [approximately 10^8 colony-forming units per milliliter (CFU/mL)] (GLI Mycobacteriology Laboratory Manual, *http://www.stoptb.org/wg/gli/assets/documents/gli_mycobacteriology_lab_manual_web.pdf*, last accessed July 23, 2021). Serial dilutions in phosphate buffer supplemented with 0.025% Tween 80 (Merck, Sandton, South Africa) were inoculated onto Middlebrook 7H10 solid media (Becton Dickinson) and incubated for 21 days at 37°C for CFU calculations. These experiments were performed in biological triplicate. One milliliter of the 10^4, 10^3, and 10^2 CFU/mL suspensions were GenoLysed (Hain Lifescience) and MTBDR*sl* performed per the manufacturer's instructions (Hain Lifescience). The two lower dilutions approximate to smear-negative disease (<10,000 CFU/mL),[11] expected to be most affected by a suboptimal ramp rate. DNA was amplified with the CFX96 thermocycler (Bio-Rad Laboratories, Sandton, South Africa) at ramp rates of 2.2°C/second and 4.0°C/second. Two experienced readers recorded bands in a blinded manner. Accuracy analyses for TUB-band positivity, indeterminate rates, incorrect banding calls, and incorrect drug class diagnostic calls were done.

Impact of Thermocycler Ramp Rate on MTBDR*sl* Performance on Clinical Specimens

GenoLysed samples ($n = 52$) remaining after programmatic line probe assay test results were collected from a TB laboratory in Cape Town, South Africa. These samples were, per the national algorithm, derived from the paired sputum specimen of a presumptive pulmonary TB patient who received Xpert MTB/RIF Ultra (Ultra) (on separate sputum), MGIT 960 culture, and Auramine O microscopy (on the same sputum before being GenoLysed). All sputa were smear-negative and Ultra-positive rifampicin-resistant.

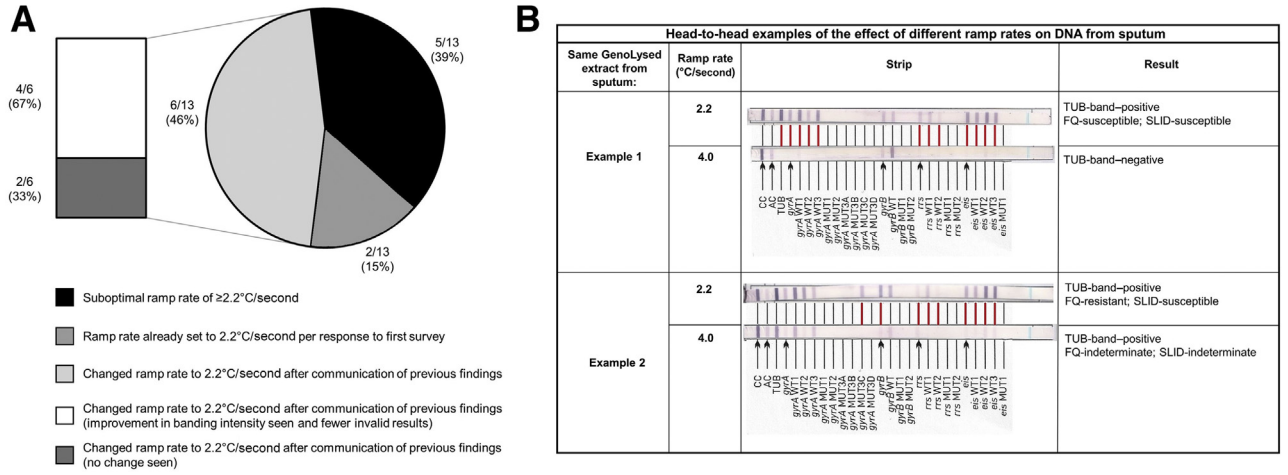


Figure 2 **A:** Follow-up survey results summarizing thermocycler ramp rates for GenoType MTBDR*sl* VER 2.0. Two (15%) of 13 initially surveyed laboratories already had their ramp rate set to 2.2°C/second, and five (39%) of 13 were still using a suboptimal ramp rate of ≥2.2°C/second upon resurveying. Six (46%) of 13 laboratories had, since the first survey on GenoType MTBDR*plus* VER 2.0, changed the GenoType MTBDR*sl* VER 2.0 ramp rate to the recommended ramp rate. Of these, four (67%) of six reported an improvement in banding intensity and fewer invalid results. **B:** An illustrative example of differences in banding patterns (and consequences for patient diagnoses) caused using suboptimal ramp rate. In example 1, at the suboptimal ramp rate (4.0°C/second), no tuberculosis or drug susceptibility information would be generated. In example 2, at the suboptimal ramp rate (4.0°C/second), again no drug susceptibility information would be generated, but, in this case, it would lead to a missed diagnosis of fluoroquinolone (FQ) resistance. Different banding patterns between strips are shown with a **red line**. SLID, second-line injectables; TUB-band—positive, positive for *Mycobacterium tuberculosis*—complex-DNA.

Smear-positive specimens were not included as it was previously shown that ramp rate had no effect on MTBDR*plus* performance on smear-positive specimens.[7] Residual GenoLysed samples were stored at −20°C.

Samples were categorized by using programmatic line probe assay results as: 17 MDR-TB, 24 pre-XDR, and 11 XDR-TB. For the experiment, DNA was amplified by using a CFX96 thermocycler (Bio-Rad, Hercules, CA) at 2.2°C/second (manufacturer-recommended) and 4.0°C/second. MTBDR*sl* was performed per the manufacturer's instructions (Hain Lifescience), and two experienced readers recorded bands in a blinded fashion. Accuracy analyses for TUB-band positivity, indeterminate rates, incorrect banding calls, and incorrect drug class diagnostic calls were done.

Calculation of Laboratory Savings from an Improvement in MTBDR*sl* Performance on Smear-Negative Specimens Stemming from Ramp Rate

Calculations were performed on how much the routine laboratory, from which GenoLysed remnants were received, would save if the proportional increase, which was found in valid results when the optimal versus the suboptimal ramp rate was used, was applied. This cost savings calculation was based on the average number of MTBDR*sl* tests performed indirectly on cultured isolates per month (which would now be reduced due to direct testing on smear-negative specimens having improved performance) and the cost of each test (including consumables, labor, and overheads; the sum is pre-calculated and supplied by the laboratory provider).

Inter-reader Agreement

An additional three experienced readers, independent of the aforementioned two readers, read all strips from the dilution series and clinical specimens at either ramp rate independently from one another and blinded to each other's calls as well as any other information regarding the specimens or strains used. Banding calls were assessed between readers, as well as resultant differences in final drug class diagnostic calls. Excluding the CC-bands and AC-bands, and including the TUB-band, gene locus control-bands, and gene-specific wild type- and mutant-bands, there are 25 possible bands per MTBDR*sl* strip. There are hence 450 possible bands total for the 18 samples in the dilution series and 1300 possible bands for the 52 clinical isolates. Each strip results in two drug class diagnostic calls, and there are hence 36 possible drug class diagnostic calls in total for 18 samples in the dilution series and 104 possible drug class diagnostic calls in total for the 52 clinical isolates.

Follow-Up Survey of TB Diagnostic and Research Laboratories

Prior respondents to the initial MTBDR*plus*-focused survey[7] were re-surveyed ($n = 29$) to gather information on the current MTBDR*sl* conditions. Other laboratories newly known to us as performing MTBDR*sl* on smear-negative specimens ($n = 11$) were also surveyed for the first time, and initial nonresponders were re-contacted at least twice. Survey questions included whether ramp rate changed and impact on nonvalid results (Supplemental Appendix S1[7]). Permission to use data in an anonymized manner was

Table 3 Comparison of Banding and Drug Class Diagnostic Calls Done on a Dilution Series of DS-TB and XDR-TB Strains and Clinical Specimens Interpreted by Three Experienced Readers

	DS-TB strain		XDR-TB strain		Clinical specimens	
Ramp rate (°C/second)	Different banding call between readers	Different drug class diagnostic call between readers	Different banding call between readers	Different drug class diagnostic call between readers	Different banding call between readers	Different drug class diagnostic call between readers
2.2	0/225* (0)	0/18† (0)	1/225* (0.4)	0/18† (0)	34/1300‡ (3)	5/104§ (5)
4.0	1/225* (0.4), $P = 0.317$	1/18† (6), $P = 0.311$	3/225* (1), $P = 0.313$	0/18† (0), $P > 0.999$	52/1300‡ (4), **$P = 0.030$**	8/104§ (8), $P = 0.390$

Data are expressed as *n/N* (%). Differences in banding calls or drug class diagnostic calls did not differ between the three readers at either ramp rate for the dilution series of cells, neither did the drug class diagnostic calls in the clinical specimens; however, significant difference between readers for banding calls on the clinical sputa occurred. *P* values are for within-column comparisons between different ramp rates. Significant *P* values are marked in bold. banding call, the presence or absence of a band deviating from the true banding call; diagnostic call, the presence or absence of banding patterns resulting in deviation of the true susceptibility to a drug class; DS-TB, drug-susceptible tuberculosis; XDR-TB, extensively drug-resistant tuberculosis.

*One strain × 3 replicates × 3 dilutions × 25 bands per strip.

†One strain × 3 replicates × 3 dilutions × 2 drug class diagnostic calls.

‡Fifty-two clinical specimens × 25 bands per strip.

§Fifty-two clinical specimens × 2 drug class diagnostic calls.

received from the Faculty of Medicine and Health Sciences Human Research Ethics Committee of Stellenbosch University (N16/04/045).

Statistical Analyses

Data were analyzed using Stata version 15 (StataCorp, College Station, TX) and GraphPad Prism version 8.0.1 (GraphPad Software, La Jolla, CA) using two-sided *t*-tests with $\alpha = 0.05$. McNemar's test was used to calculate differences for paired data (ie, the same DNA tested at both ramp rates). The two-sample proportion test was used for comparisons between proportions.

Results

MTBDR*sl* on the Dilution Series at Different Ramp Rates

Overall, there were no differences between ramp rates of 2.2°C/second and 4.0°C/second for TUB-band detection [16 of 18 (89%) versus 17 of 18 (94%); $P = 0.547$], indeterminate results [2 of 16 (13%) versus 3 of 17 (18%); $P = 0.680$], incorrect banding calls [22 of 400 (6%) versus 33 of 425 (8%); $P = 0.193$)], or incorrect drug resistance calls [2 of 32 (6%) versus 2 of 34 (6%); $P = 0.950$] (Table 1). Therefore, valid results did not differ significantly [14 of 16 (88%) versus 14 of 17 (82%); $P = 0.680$].

MTBDR*sl* on Clinical Sputa at Different Ramp Rates

No TUB-band detection differences were seen at 2.2°C/second versus 4.0°C/second [52 of 52 (100%) versus 51 of 52 (98%); $P = 0.315$; one MDR-TB patient was TUB-negative only at 4.0°C/second]. However, indeterminate rates improved at 2.2°C/second [0 of 52 (0%) versus 7 of 51 (14%); $P = 0.006$], as did the proportion of bands that appeared incorrectly [0 of 1300 (0%) versus 55 of 1275 (4%); $P < 0.001$)] and drug-resistance calls [0 of 104 (0%) versus 6 of 102 (6%); $P = 0.013$] (Table 2). The proportion of patients with a valid result was therefore 52 (100%) of 52 versus 41 (80%) of 51. In other words, the patients who successfully received testing for FQs and SLIDs thus improved 21% (95% CI, 8–34; $P < 0.001$).

Programmatic Ultra semi-quantitative data were available for 41 (79%) of 52 sputa. When bacterial load in sputa that gave a valid result at 2.2°C/second was compared versus sputa that gave a valid result at 4.0°C/second, there were no differences [median (interquartile range) minimum cycle threshold (C_{Tmin}), 18.7 (17.7–19.9) versus 18.8 (18.0–19.9); $P = 0.899$]. It was expected that 2.2°C/second would result in an improved limit of detection in MTBDR*sl* (better ability to detect higher C_{Tmin} and therefore fewer bacilli); however, no differences were detected.

Head-to-head examples of the effect of different ramp rates on DNA from sputum are provided in Figure 2B.

Banding patterns from both the dilution series and clinical sputa are listed in Supplemental Tables S1 and S2. For the dilution series (Supplemental Table S1), irrespective of ramp rate, MTBDR*sl* did not classify the XDR-TB strain correctly at 10^2 CFU/mL across all replicates (Table 1). Overall, for dilution series (both strains, all dilutions), the overall effect was missed resistance due to a TUB-negative, indeterminate, or a missing gene-specific band, or false-resistance due to an erroneously absent wild-type band. For clinical sputa (Supplemental Table S2) at the suboptimal ramp rate, there was worse detection of the TB and locus control bands and, when TB was detected and the locus control bands present, gene-specific bands that should have been present were absent. In the dilution series, one replicate (XDR-TB, 10 to 2 dilution) missed amikacin resistance at the suboptimal ramp rate. In clinical specimens, two samples (RR2-31 and RR2-38) with high-level moxifloxacin

Table 4 Laboratories That Indicated Their Ramp Rate Had Not yet Changed to the Manufacturer-Recommended Ramp Rate of ≤2.2°C/second Since the Last Survey, the Reason Why, and Total Number of Line Probe Assays Performed per Month

Country	Reason given	No. of line probe assays performed per month by this respondent laboratory
Kenya	Do not know	240
South Africa	Ramp rate change was not necessary as MTBDR*plus* assays are performed on cultured isolates only and no MTBDR*sl* assays are performed, as well as any changes to a standard operating procedure requires a validation process	40
Belarus	Ramp rate change in a standard operation procedure is not permitted without a prior approval process	155
Denmark	Ramp rate was not changed due to the run time of the original amplification protocol being faster	25
Spain	The thermocycler did not permit a ramp rate change	12

These laboratories perform either GenoType MTBDR*plus* VER 2.0 (MTBDR*plus*), GenoType MTBDR*sl* VER 2.0 (MTBDR*sl*), or both on smear-negative specimens, but data on the subtotals for each assay were not collected.

resistance were incorrectly classified at the suboptimal ramp rate as low-level resistant (RR2-38) or susceptible (RR2-31). At the suboptimal ramp rate of 4.0°C/second, 55 gene locus bands were erroneous. The breakdown is as follows: *gyrA*, 14 of 55 (25%); *gyrB*, 5 of 55 (9%); *rrs*, 28 of 55 (51%); and *eis*, 8 of 55 (15%).

Laboratory Savings

If the improvement in FQ and SLID testing due to optimal ramp rate usage is applied, there would be a 21% decrease in the number of tests required to be performed indirectly (which would require culture and a second MTBDR*sl*). At a local reference laboratory, approximately 320 MTBDR*sl* assays, initially attempted on smear-negative sputa, are performed per month and are subsequently repeated on culture isolates. Hence, in a scenario in which this laboratory was using an incorrect ramp rate and changed to the correct rate, they would perform approximately 67 fewer indirect MTBDR*sl* assays per month. At a total per test cost of US$60 (6% per annum inflation),[12] this translates to a savings of US$48,240 per year (only factoring in pure laboratory costs).

Inter-reader Agreement

In the dilution series, diagnostic calls did not differ between the three readers at either ramp rate. All readers incorrectly classified the XDR-TB strain (as either TUB-band−negative or indeterminate) at all 10^2 CFU/mL replicates and the drug-susceptible−TB strain (as indeterminate) at one of the three replicates at 10^2 CFU/mL (Table 3). The proportion of disagreement between readers (banding calls) did not differ at suboptimal versus optimal ramp rates [for the drug-susceptible (1 of 225 versus 0 of 225; $P = 0.317$) or the XDR (3 of 225 versus 1 of 225; $P = 0.313$)] strain.

In clinical sputa, however, although the disagreement in drug class diagnostic calls did not differ between readers at the optimal versus suboptimal ramp rate [5 of 104 (5%) versus 8 of 104 (8%); $P = 0.390$], banding calls did differ [34 of 1300 (3%) versus 52 of 1300 (4%); $P = 0.030$].

Additional Survey

Twenty-nine follow-up surveys were sent to the original respondents and 11 to new laboratories. Thirteen total responses were received (45%), including four from new respondents (Figure 2A). Two (15%) of 13 respondents already had their ramp rate at 2.2°C/second (per their response to the first survey), and six (46%) of 13 had subsequently changed their ramp rate to 2.2°C/second after the previous findings were communicated.[7] Concerningly, five (39%) of 13 had not changed, for which varied reasons were offered (Table 4). Of the laboratories who changed to 2.2°C/second, four (67%) of six reported that this resulted in an improvement in banding intensity and fewer nonvalid results for MTBDR*plus* and MTBDR*sl*.

Discussion

The current study evaluated for the first-time the impact of thermocycler ramp rates on the most widely used molecular test for second-line drug-resistant TB (MTBDR*sl*). This study shows: i) in sputa, valid results improved by 21% when using the optimal ramp rate, which results in significant laboratory cost savings and would decrease diagnostic delay; ii) banding call and drug susceptibility call reader disagreement worsened at the suboptimal ramp rate; and iii) several laboratory respondents had not corrected their line probe assay ramp rate but, those that had, reported fewer nonvalid results from MTBDR*sl* on smear-negative specimens.

In a previous study, the authors found that a suboptimal thermocycler ramp rate negatively affects the diagnostic accuracy of potentially thousands of MTBDR*plus* assays, especially on smear-negative sputa,[7] and ramp rate

monitoring was incorporated into laboratory quality control and training documentation (WHO Drug-resistant tuberculosis: how to interpret rapid molecular test results, *https://openwho.org/courses/multi-drug-resistant-tb*, last accessed July 6, 2021). The current study shows that a 21% increase in MTBDR*sl* diagnoses (valid results) in smear-negative specimens is possible through ramp rate correction. This is not a niche problem; diagnostic laboratories that still do not perform MTBDR*sl* correctly were identified. This correction, which this study has now provided MTBDR*sl*-specific empirical evidence, could reduce drug-resistant TB diagnostic care cascade gaps: a recent study found that only 65% of MDR-TB cases were evaluated for FQ resistance.[13]

Critically, ramp rate correction will reduce repeat MTBDR*sl* testing on isolates. Most directly, this will translate into substantial laboratory cost savings in high-burden countries, especially when TB services are fragile due to the COVID-19 pandemic, not to mention the myriad of other individual and population benefits that can stem from improved drug susceptibility testing[14]; these include reduced time to treatment, transmission, and mortality.

Most laboratories in the follow-up survey had corrected the ramp rate; however, a significant amount, including those responsible for routine diagnostic testing on smear-negative specimens, still used a suboptimal ramp rate. It should be emphasized that: i) laboratories must ensure that they are using the optimal ramp rate; ii) thermocycler ramp rate monitoring should be added to laboratory external quality assurance programs and accreditation processes for MTBDR*sl*; and iii) the manufacturer should make the recommended ramp rate more prominent in assay documentation. It is worth evaluating further why incorrect ramp rates continued to be used. This may be due to quality assurance lapses, a deliberate choice (eg, to potentially speed up turn-around-time) without an awareness of downsides, or a design limitation of available thermocyclers.

When a band was present at the optimal ramp rate (2.2°C/second) and not the suboptimal ramp rate (4.0°C/second), FQ and/or SLID diagnoses were missed completely due to gene locus control bands not binding. False drug class diagnostic calls for FQs and/or SLIDs (false resistance) due to the inability of a band to bind were also seen. No false resistance was observed due to the binding of mutant probes when the suboptimal ramp rate was used. However, false resistance calls due to an erroneous absence of wild-type bands occurred. It was noted that more than one-half of the incorrect bands in sputa occurred in one gene locus (*rrs*), which may be due to secondary structures that interfere with PCR and detection.

A more prominent performance difference was seen between ramp rates in clinical sputa than in spiked solution. Bacilli in mucus sputa matrices behave differently from bacilli spiked in *in vitro* experiments, and these findings illustrate potential downsides to investigating the effect of PCR parameters on molecular assays when *in vitro* or mock specimens are used.

The current evaluation has strengths and limitations. A wider ramp rate range or different thermocycler models were not assessed due to limited sputa and cost. The utility of additional testing when a useful (ie, valid) result failed to be generated was also not evaluated. The most frequently reported incorrect ramp rate from the previous survey was used.[7] DNA from samples was not directly quantified; however, when comparing Ultra semi-quantitative (C_{Tmin}) data between valid results across ramp rates, no differences occurred. When there is an indeterminate result for a gene locus, regardless of whether that indeterminate result is caused by optimal ramp rate, it may influence the reliability of other diagnostic calls from loci with valid control bands. However, this requires a large diagnostic accuracy study to investigate, and the current work was not designed to do so.

The survey results would have also been subjected to selection, response, and reporting biases. The authors suggest that a formal survey be done by the manufacturer and/or the appropriate regulatory and oversight agency (the study survey was done independently). Savings stemming from quicker diagnosis, treatment initiation, and long-term reductions in transmission and mortality due to improved performance were not evaluated; there is already a saving in laboratory costs alone, with no downside.

In conclusion, this study found that a still incorrectly configured and innocuous technical setting (ramp rate) has a real-world negative impact on patients' diagnoses for second-line drug resistance using MTBDR*sl*. Patients with smear-negative specimens, for whom early diagnosis is important to curtail transmission of drug resistance, are especially vulnerable. All stakeholders must ensure that the optimal thermocycler ramp rate for MTBDR*sl* is used, and the impact of this source of technical variation should be investigated for other molecular diagnostics.

Acknowledgments

We thank the laboratories that participated in the survey and provided data, the National Health Laboratory Service (Green Point, Cape Town, South Africa), and Hain Lifescience, which donated the MTBDR*sl* kits used for the dilution series.

Author Contributions

B.D., M.d.V., G.T., and R.W. conceived the experiments; T.D., N.B., and S.P. provided specimens and data from the National Health Laboratory Service; B.D. conducted the experiments and analyzed the data; S.P, Y.G., R.V., and S.M. assisted with analysis of results; J.M. provided critical input. All authors reviewed the manuscript.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.jmoldx.2022.01.003*.

References

1. World Health Organization: Global Tuberculosis Report 2020. Geneva, Switzerland, WHO, 2020. Available at: https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf (accessed July 6, 2021)
2. Basu S, Friedland GH, Medlock J, Andrews JR, Shah NS, Gandhi NR, Moll A, Moodley P, Sturm AW, Galvani AP: Averting epidemics of extensively drug-resistant tuberculosis. Proc Natl Acad Sci U S A 2009, 106:7672−7677
3. Dorman SE, Nahid P, Kurbatova EV, Phillips PPJ, Bryant K, Dooley KE, Engle M, Goldberg SV, Phan HTT, Hakim J, Johnson JL, Lourens M, Martinson NA, Muzanyi G, Narunsky K, Nerette S, Nguyen NV, Pham TH, Pierre S, Purfield AE, Samaneka W, Savic RM, Sanne I, Scott NA, Shenje J, Sizemore E, Vernon A, Waja Z, Weiner M, Swindells S, Chaisson RE; AIDS Clinical Trials Group; Tuberculosis Trials Consortium: Four-month rifapentine regimens with or without moxifloxacin for tuberculosis. N Engl J Med 2021, 384:1705−1718
4. World Health Organization: The use of molecular line probe assays for the detection of resistance to second-line anti-tuberculosis drugs: policy guidance. Geneva, Switzerland, WHO, 2016. Available at: https://apps.who.int/iris/handle/10665/246131 (accessed July 6, 2021)
5. World Health Organization: WHO consolidated guidelines on tuberculosis. Module 3: Diagnosis—Rapid diagnostics for tuberculosis detection 2021 update. Geneva, Switzerland, WHO, 2021. Available at: https://www.who.int/publications/i/item/9789240029415 (accessed July 6, 2021)
6. Theron G, Peter J, Richardson M, Warren R, Dheda K, Steingart KR: GenoType® MTBDRsl assay for resistance to second-line anti-tuberculosis drugs. Cochrane Database Syst Rev 2016, 9:CD010705
7. Derendinger B, de Vos M, Nathavitharana RR, Dolby T, Simpson JA, van Helden PD, Warren RM, Theron G: Widespread use of incorrect PCR ramp rate negatively impacts multidrug-resistant tuberculosis diagnosis (MTBDRplus). Sci Rep 2018, 8:3206
8. Dowdy DW, Theron G, Tornheim JA, Warren R, Kendall EA: Of testing and treatment: implications of implementing new regimens for multidrug-resistant tuberculosis. Clin Infect Dis 2017, 65:1206−1211
9. World Health Organization: WHO consolidated guidelines on drug-resistant tuberculosis treatment. Geneva, Switzerland, WHO, 2019. Available at: https://apps.who.int/iris/handle/10665/311389 (accessed July 6, 2021)
10. World Health Organization: Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance. Geneva, Switzerland, WHO, 2021. Available at: https://www.who.int/publications/i/item/9789240028173 (accessed July 6, 2021)
11. Hobby GL, Holman AP, Iseman MD, Jones JM: Enumeration of tubercle bacilli in sputum of patients with pulmonary tuberculosis. Antimicrob Agents Chemother 1973, 4:94−104
12. Groessl EJ, Ganiats TG, Hillery N, Trollip A, Jackson RL, Catanzaro DG, Rodwell TC, Garfein RS, Rodrigues C, Crudu V, Victor TC, Catanzaro A: Cost analysis of rapid diagnostics for drug-resistant tuberculosis. BMC Infect Dis 2018, 18:102
13. De Vos E, Scott L, Voss De Lima Y, Warren RM, Stevens W, Hayes C, da Silva P, Van Rie A: Management of rifampicin-resistant TB: programme indicators and care cascade analysis in South Africa. Int J Tuberc Lung Dis 2021, 25:134−141
14. Dheda K, Gumbo T, Maartens G, Dooley KE, Murray M, Furin J, Nardell EA, Warren RM; Lancet Respiratory Medicine drug-resistant tuberculosis Commission group: The Lancet Respiratory Medicine Commission: 2019 update: epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant and incurable tuberculosis. Lancet Respir Med 2019, 7:820−826

Appendix VII

More than Mycobacterium tuberculosis: site-of-disease microbial communities, and their functional and clinical profiles in tuberculous lymphadenitis

Publication status: published

Nyawo, G., Naidoo, C., Wu, B., Sulaiman, I., Clemente, J., Li, Y., **Minnies, S.**, Reeve, B., Moodley, S., Rautenbach, C., Wright, C., Singh, S., Whitelaw, A., Schubert, P., Warren, R., Segal, L., Theron, G. More than Mycobacterium tuberculosis: site-of-disease microbial communities, and their functional and clinical profiles in tuberculous lymphadenitis. *Thorax*, *78*(3), pp.297-308.

DOI: http://dx.doi.org/10.1136/thorax-2022-219103

OPEN ACCESS

Original research

More than *Mycobacterium tuberculosis:* site-of-disease microbial communities, and their functional and clinical profiles in tuberculous lymphadenitis

Georgina R Nyawo,[1,2] Charissa C Naidoo,[1,2] Benjamin Wu,[3] Imran Sulaiman,[3] Jose C Clemente,[4] Yonghua Li,[3] Stephanie Minnies,[1] Byron W P Reeve,[1] Suventha Moodley,[1,2] Cornelia Rautenbach,[5,6] Colleen Wright,[7] Shivani Singh,[3] Andrew Whitelaw,[5,6] Pawel Schubert,[5,7] Robin Warren,[1] Leopoldo Segal,[3] Grant Theron [1,2]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/thorax-2022-219103).

For numbered affiliations see end of article.

Correspondence to
Professor Grant Theron, DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Stellenbosch University Faculty of Medicine and Health Sciences, Cape Town, Western Cape, South Africa; gtheron@sun.ac.za

TBScience 2022, hosted at the Union World Conference on Lung Health 2022.

Received 20 April 2022
Accepted 31 October 2022
Published Online First
8 December 2022

To cite: Nyawo GR, Naidoo CC, Wu B, *et al*. *Thorax* 2023;**78**:297–308.

ABSTRACT

Background Lymphadenitis is the most common extrapulmonary tuberculosis (EPTB) manifestation. The microbiome is important to human health but uninvestigated in EPTB. We profiled the site-of-disease lymph node microbiome in tuberculosis lymphadenitis (TBL).

Methods Fine-needle aspiration biopsies were collected from 158 pretreatment presumptive TBL patients in Cape Town, South Africa. 16S Illumina MiSeq rRNA gene sequencing was done.

Results We analysed 89 definite TBLs (dTBLs) and 61 non-TBLs (nTBLs), which had similar α- but different β-diversities (p=*0.001*). Clustering identified five lymphotypes prior to TB status stratification: *Mycobacterium*-dominant, *Prevotella*-dominant and *Streptococcus*-dominant lymphotypes were more frequent in dTBLs whereas a *Corynebacterium*-dominant lymphotype and a fifth lymphotype (no dominant taxon) were more frequent in nTBLs. When restricted to dTBLs, clustering identified a *Mycobacterium*-dominant lymphotype with low α-diversity and non-*Mycobacterium*-dominated lymphotypes (termed *Prevotella-Corynebacterium*, *Prevotella-Streptococcus*). The *Mycobacterium* dTBL lymphotype was associated with HIV-positivity and features characteristic of severe lymphadenitis (eg, larger nodes). dTBL microbial communities were enriched with potentially proinflammatory microbial short-chain fatty acid metabolic pathways (propanoate, butanoate) vs nTBLs. 11% (7/61) of nTBLs had *Mycobacterium* reads BLAST-confirmed as *Mycobacterium tuberculosis* complex.

Conclusions TBL at the site-of-disease is not microbially homogeneous. Distinct microbial community clusters exist that, in our setting, are associated with different clinical characteristics, and immunomodulatory potentials. Non-*Mycobacterium*-dominated dTBL lymphotypes, which contain taxa potentially targeted by TB treatment, were associated with milder, potentially earlier stage disease. These investigations lay foundations for studying the microbiome's role in lymphatic TB. The long-term clinical significance of these lymphotypes requires prospective validation.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Lymphadenitis is the most frequent extrapulmonary tuberculosis manifestation. The microbiome is critical for human health, however, the microbiome at the site-of-disease in patients with tuberculosis lymphadenitis is completely uncharacterised, including whether distinct microbial clusters (which we term 'lymphotypes') are associated with clinically important patient characteristics.

WHAT THIS STUDY ADDS

⇒ Surprisingly, patients with confirmed tuberculosis lymphadenitis often had bacterial taxa other than *Mycobacterium* dominant at the site-of-disease (*Prevotella, Streptococcus, Corynebacterium*). Such patients had milder forms of disease (eg, less swelling, less HIV) whereas patients with the *Mycobacterium*-dominated lymphotype had increased microbial functional capacity for proinflammatory short-chain fatty acids and more severe disease.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our findings have relevance for clinical staging and treatment of tuberculosis lymphadenitis, which we show to not be microbially homogeneous, and suggest that the site-of-disease in tuberculosis lymphadenitis is, prior to shifting to becoming *Mycobacterium*-dominated, first characterised by *Prevotella, Streptococcus* and/or *Corynebacterium* dominance and milder disease. Lastly, given that *Streptococcus* and *Corynebacterium* are themselves capable of causing lymphadenitis and susceptible to first-line TB treatment, such treatment may alleviate pathology in tuberculosis lymphadenitis by, in part, killing taxa other than *Mycobacterium*.

INTRODUCTION

Tuberculosis (TB), which kills 1.5 million people globally each year (including 214 000 people

with HIV), causes extrapulmonary TB (EPTB).[1] EPTB accounts for ~16% of all TB, up to half of all TB in people living with HIV (PLHIV)[2] and has high mortality.

TB lymphadenitis (TBL) is the most common EPTB manifestation, accounting for 70% of EPTB and most frequently affects peripheral and cervical lymph nodes.[3 4] TBL occurs after *Mycobacterium tuberculosis* (*Mtb*) enters the airways, is taken up by phagocytic cells, and transported to lymph nodes where granulomas may form. These steps are also necessary for priming T-cells to generate adaptive immune responses for microbial killing mediated by cytokines and other effector mechanisms.[5]

Lymph nodes have an important role in TB pathogenesis: enlargement has been documented following exposure, even if only a fraction of patients with enlarged nodes develop active disease.[6] Animal studies show lymph nodes can be sites of TB reactivation (*Mtb* DNA found in new lung granulomas share unique DNA barcodes with *Mtb* previously only found in lymph nodes).[7] Furthermore, pathologically normal lymph nodes obtained at autopsy from humans without active TB can, when used to inoculate animals, cause active disease,[8] suggesting these lymph nodes contained live *Mtb* (and hence *Mtb* DNA). Lymph nodes are therefore hypothesised to serve as a *Mtb* growth and persistence niche[6] that can spread to bodily sites[9] (in animals lymph node infection almost always accompanies infection in the lungs[7]; suggesting that TB may primarily be a lymphatic rather than pulmonary disease.[10] For example, the lymph nodes of participants with subclinical TB pathology demonstrate enhanced metabolic activity on positron emission tomography (PET)-CT scans.[11] Together these studies show that lymph nodes have an important role in TB pathogenesis, however, the determinants of why *Mtb* sometimes successfully establishes itself in the lymph nodes and subsequently proliferates, including the potential role of other microbes, is understudied. Key to understanding this is characterising the local site-of-disease.

The microbiota modulates immune responses via microbially-derived metabolites known as short-chain fatty acids (SCFAs).[12] Enriched pulmonary SCFAs predict TB risk in HIV-infected individuals stable on ART, and ex vivo addition of butyrate inhibits *Mtb*-induced proinflammatory responses.[13] Two studies assessed lymph node microbial content,[14 15] both in mesenteric lymph nodes in Crohn's disease where reduced diversity was observed. The site-of-disease microbiome in TB is underexamined[16]: in bronchoalveolar lavage fluid (BALF), active pulmonary TB was associated with *Mycobacterium* enrichment and *Streptococcus* depletion.[17 18]

The site-of-disease microbiome in TBL (including in HIV-endemic settings where TB is common) remains uncharacterised. Therefore, given the apparent role of the lymph nodes in TB pathogenesis, and the importance of the microbiome as a modulator of immunity, we characterised the site-of-disease lymph microbiome in presumptive TBL patients from a high HIV burden setting[19] before the potentially confounding effects of antibiotic-based TB treatment.

METHODS

Patient recruitment and follow-up

Presumptive TBL participants (≥18 years) were recruited from Tygerberg Academic Hospital in Cape Town, SA (25 January 2017–11 December 2018). Participants were programmatically referred for a routine fine needle aspiration biopsy (FNAB) via the skin for the investigation of lymphadenopathy as described.[19] Eligible participants were not on TB treatment within 6 months. Clinical and demographic data were collected by interview and medical record review. Patients programmatically diagnosed with TBLwere initiated on treatment, and study staff assessed treatment response by telephonic follow-up ≥12 weeks. The study had no role in patient management.

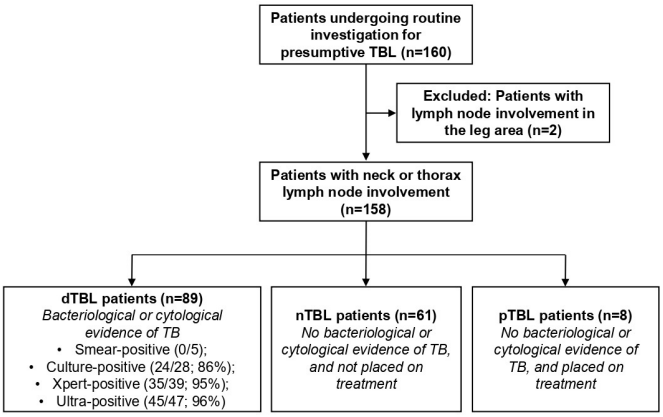


Figure 1 Study flow chart. Fine-needle aspirates, skin and saline controls were collected from presumptive TBL patients. dTBLs, definite-TBL; MGIT960 culture, mycobacteria growth indicator tube 960 liquid culture; nTBLs, non-TBLs; pTBLs, probable TBLs; Smear: Smear microscopy; Ultra: Xpert MTB/RIF Ultra; Xpert: Xpert MTB/RIF.

Specimen collection and processing

For each patient, two background DNA sampling controls were collected in microcentrifuge tubes prior to lymph node aspiration: a skin swab (collected into saline; Ysterplaat Medical Supplies, Cape Town, South Africa) of the site to be punctured, followed by a saline flush of the syringe to be used for aspiration. Aspiration and microbiological procedures are in online supplemental methods. Aspirated material from the third pass was collected into 500 µL sterile saline and stored at −80°C until batched DNA extraction.

Routine specimen testing

Patients were categorised based on lymphatic or non-lymphatic mycobacteriological evidence, provided by the government programmatic laboratory (National Health Laboratory Service), and/or clinical decision to start treatment by the responsible clinician thereafter.

Case definitions

Briefly, definite-TBLs (dTBLs) had at least one *Mtb* complex (MTBC)-positive extrapulmonary or pulmonary specimen by Xpert or culture (figure 1). Alternatively, they had site-of-disease cytology compatible with active TB. Probable-TBLs (pTBLs) did not meet dTBL criteria but commenced treatment empirically. Non-TBLs (nTBLs) had no microbiological or cytological evidence of TB. Further detail is in online supplemental table S1.

Microbial DNA extraction and sequencing

DNA was extracted from specimens and controls using the PureLink Microbiome DNA Purification Kit (Invitrogen, Carlsbad, USA). The 16S rRNA gene V4 hypervariable region (150 bp read length) was amplified and sequenced

(paired-ends) on the Illumina MiSeq platform. Lymph, skin swab and one in five saline flushes were extracted and sequenced.

Microbiome data analysis

16S rRNA gene sequences were processed, denoised and analysed in Quantitative Insights Into Microbial Ecology (QIIME 2, v2020.8)[20] and DADA2[21] using closed-reference picking by assigning taxonomy at a 97% similarity against representative sequences in Greengenes (V.13.8).[22] QIIME2 outputs (phylogenetic tree, feature table, taxonomy) and metadata were imported into R (V.3.5.2) and analyses done using *phyloseq*.[23] Shannon's index was calculated with *vegan*[24] as measure α-diversity (within-sample diversity). Bray-Curtis distances were calculated as a measure of β-diversity (between-sample diversity) and were visualised as principal coordinate analysis plots. Dirichlet-Multinomial Mixtures (DMM) modelling was done to estimate the optimal number of clusters based on microbial compositional similarity.[25] These clusters are herewith referred to as 'lymphotypes'.

Inferred metagenome

Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) V.2.1.3-b[26] was used to predict gene family abundance with PICRUSt2 default options (picrust2_pipeline.py). The resulting gene table was mapped against the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database, and pathway abundances were inferred from predicted KEGG ORTHOLOGY (KO) abundances.

Differential abundance analysis

Differentially abundant taxa and pathways were identified using *DESeq2* (V.1.22.2), which internally corrects and normalises data.[27] Feature tables were pruned to have a mean relative abundance ≥5% in 0.5% of samples.[28] *DESeq2* was run on PICRUSt outputs to identify common pathways in oL4 versus each lymphotype (overall patients), and dL3 versus each other lymphotype (dTBLs). *DESeq2* outputs with abundances and significance values for each discriminatory taxon and pathway were obtained (see online supplemental material: *DESeq2* Tables). A false discovery rate (FDR)-adjusted p≤0.2 and ≤0.05 was considered significant for taxa and pathways, respectively.

Statistical analyses

Statistical analysis was done in GraphPad Prism V.7 (GraphPad Software, USA), STATA V.16 (StataCorp) and R V.4.2 (R Core Team, 2022). The proportions test was done to determine whether a specific variable was more frequent in different groups (eg, patients of different TB status).[29] For analysis of microbiome data, non-parametric tests were used as microbiome data are not normally distributed.[30] The Mann-Whitney or Wilcoxon signed rank test was used for unpaired and paired comparisons between two groups respectively (eg, α-diversity). Kruskal-Wallis with Dunn's test was used for comparison involving more than two groups (eg, relative abundance comparisons). Spearman's rank correlation was used to measure the association between mycobacterial relative abundance and continuous variables (eg, lymph node size). Permutational multivariate analysis of variance (PERMANOVA) was computed with 999 permutations for β-diversity differences, and R^2 used to measure the proportion variation explained by a variable.[20] The Benjamini-Hochberg procedure was used to correct for multiple comparisons by controlling for FDR.[28] For analysis of continuous variables in different groups (eg, lymph node size different TB status), the D'Agostino-Pearson omnibus normality test was done to evaluate normality, and the relevant parametric or nonparametric test was chosen based on the normality test. A p≤0.05 was considered significant for all comparisons, unless otherwise specified.

Table 1 Demographic and clinical characteristics of patients with presumptive TBL

	Patients with presumptive TB (n=158)			
	Total (n=158)*	dTBL (n=89)	nTBL (n=61)	P value
Age, years	36 (21–44)	35 (29–40)	38 (30–49)	0.053
Female	85/159 (53)	48/89 (54)	35/61 (57)	0.677
HIV†	77/156 (49)	49/89 (55)	23/59 (39)	0.055
CD4+cells/μL	166 (90–308)	155 (76–251)	250 (139–458)	**0.027**
CD4+<200 cells/μL	47/77 (61)	32/49 (65)	11/23 (48)	0.159
On ART†	38/76 (50)	21/49 (43)	14/22 (64)	0.105
Previous TB†	36/156 (23)	24/88 (27)	9/60 (15)	0.078
Pulmonary TB	27/36 (75)	17/24 (71)	8/9 (89)	0.281
Extrapulmonary TB	9/36 (25)	7/24 (29)	1/9 (11)	0.281
Tobacco smoking†	44/157 (28)	21/89 (24)	22/60 (37)	0.084
Antibiotic use within 1 year of recruitment†	41/155 (26)	22/87 (25)	16/60 (27)	0.851
At recruitment	24/41 (59)	10/22 (45)	11/16 (69)	0.154
Lymph node characteristics: sites				
Neck	138/158 (87)	78/89 (88)	55/61 (90)	0.632
Deep anterior cervical	65/138 (47)	36/78 (46)	24/55 (47)	0.774
Deep lateral cervical	25/138 (18)	15/78 (19)	10/55 (18)	0.879
Superficial	15/138 (11)	6/78 (8)	9/55 (16)	0.120
Supraclavicular	19/138 (14)	16/78 (21)	3/55 (5)	**0.015**
Head	13/138 (10)	4/78 (5)	9/55 (16)	**0.032**
Thorax	20/158 (13)	11/89 (12)	6/61 (10)	0.632
Axillary (vs breast)	16/21 (81)	9/11 (82)	3/5 (60)	0.350
Lymph node characteristics: size, cm^2	4 (2–9)	4 (2–9)	4 (4–9)	0.150
Specimen appearance				
Bloody (vs chylous)	130/158 (82)	66/89 (72)	57/61 (93)	**0.003**

Bolded items indicate that p values are significant at p<0.05.
Pulmonary or extrapulmonary previous TB refers to the most recent prior TB episode.
Data are n/N (%) or median (IQR).
*Probable TBLs (pTBLs) excluded from table.
†Missing data: HIV (n=2); On ART (n=1); previous TB (n=2); smoking (n=1); antibiotic use within 1 year of recruitment (n=3).
ART, antiretroviral therapy; dTBLs, definite tuberculous lymphadenitis; nTBLs, non tuberculous lymphadenitis; pTBLs, probable-TBLs.

RESULTS

Cohort characteristics

We had 89 dTBLs, 61 nTBLs (figure 1) and 8 pTBLs (henceforth excluded due to small n), the characteristics of which are in table 1.

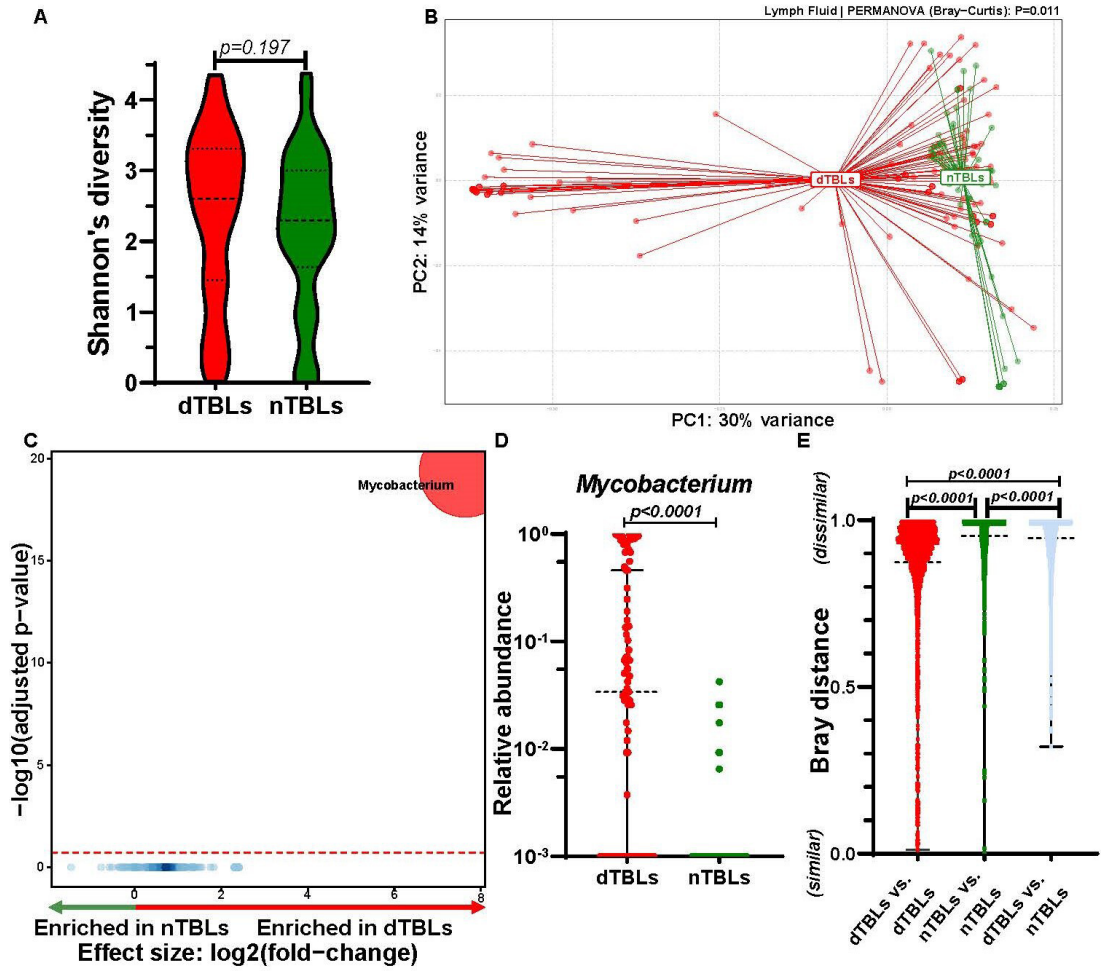


Figure 2 dTBLs have a distinct microbiome to nTBLs with *Mycobacterium* enrichment. (A) Although α-diversity was similar, (B) β-diversity differed. *Mycobacterium* was enriched in dTBLs compared with nTBLs based on (C) differential abundance testing and (D) relative abundance. Discriminatory taxa appear above the threshold (red dotted line, FDR=0.2). (E) dTBLs were more compositionally similar to each other than nTBLs. dTBLs, definite-TBLs; FDR, false discovery rate; nTBLs: non-TBL; PERMANOVA, permutational multivariate analysis of variance; TBL, tuberculous lymphadenitis.

Lymph microbiome is distinct from background sampling controls

To assess the degree of potential carry-over from skin commensals and background DNA on equipment used for biopsies, two background DNA sampling controls were collected and subjected to the same procedures as the actual samples. We compared the microbiome in skin and saline flushes with that of the lymph fluid. Lymph fluid had similar α-diversity to background controls but different β-diversity resulting from an enrichment of *Mycobacterium* (online supplemental figure S1A–D), thus background contamination is unlikely.

Mycobacterium enrichment in dTBLs drives differences with nTBLs

We evaluated the overall difference between the microbial communities of TB groups by comparing their microbial diversity and composition. α-Diversity was similar in dTBLs and nTBLs (figure 2A), but β-diversity differed and *Mycobacterium* was the most discriminatory taxon (figure 2B,C; online supplemental figure S2 has similar comparisons with pTBLs included) appearing at several fold higher frequencies than in nTBLs (figure 2D). When patients with antibiotic use within the last year were excluded, α-diversity differences by TB status were detected (lower in dTBLs, online supplemental results). Bray distances within nTBLs were greater than within dTBLs (figure 2E), thus dTBLs were more like each other than nTBLs to each other (likely reflecting the mixture of different disease pathologies in the nTBLs and relative homogeneity of dTBLs). These results show that lymph microbial communities differ in dTBLs and nTBLs, and the microbiome of TBL is characterised by a significant enrichment of *Mycobacterium*.

MTBC DNA found in tuberculous and nontuberculous lymph nodes

Mycobacterium reads were present in 64% (57/89) of dTBLs and 11% (7/61; $p<0.0001$) of nTBLs (online supplemental figure S3) and, when sequences underwent BLAST, all reads matched with *Mtb*, suggesting that none of these patients had environmental mycobacteria. There was a higher relative abundance of *Mycobacterium* reads in dTBLs (0.034 (IQR 0.001–0.460) vs 0.001 (0.001–0.001), $p<0.0001$; figure 2D), and the 16S rRNA gene sequencing positively correlated with TB diagnostic tests, but not with lymph node size (online supplemental figure S4A,B). These results suggest that MTBC DNA is found in most dTBL lymph nodes and occasionally occurs in nTBL lymph nodes.

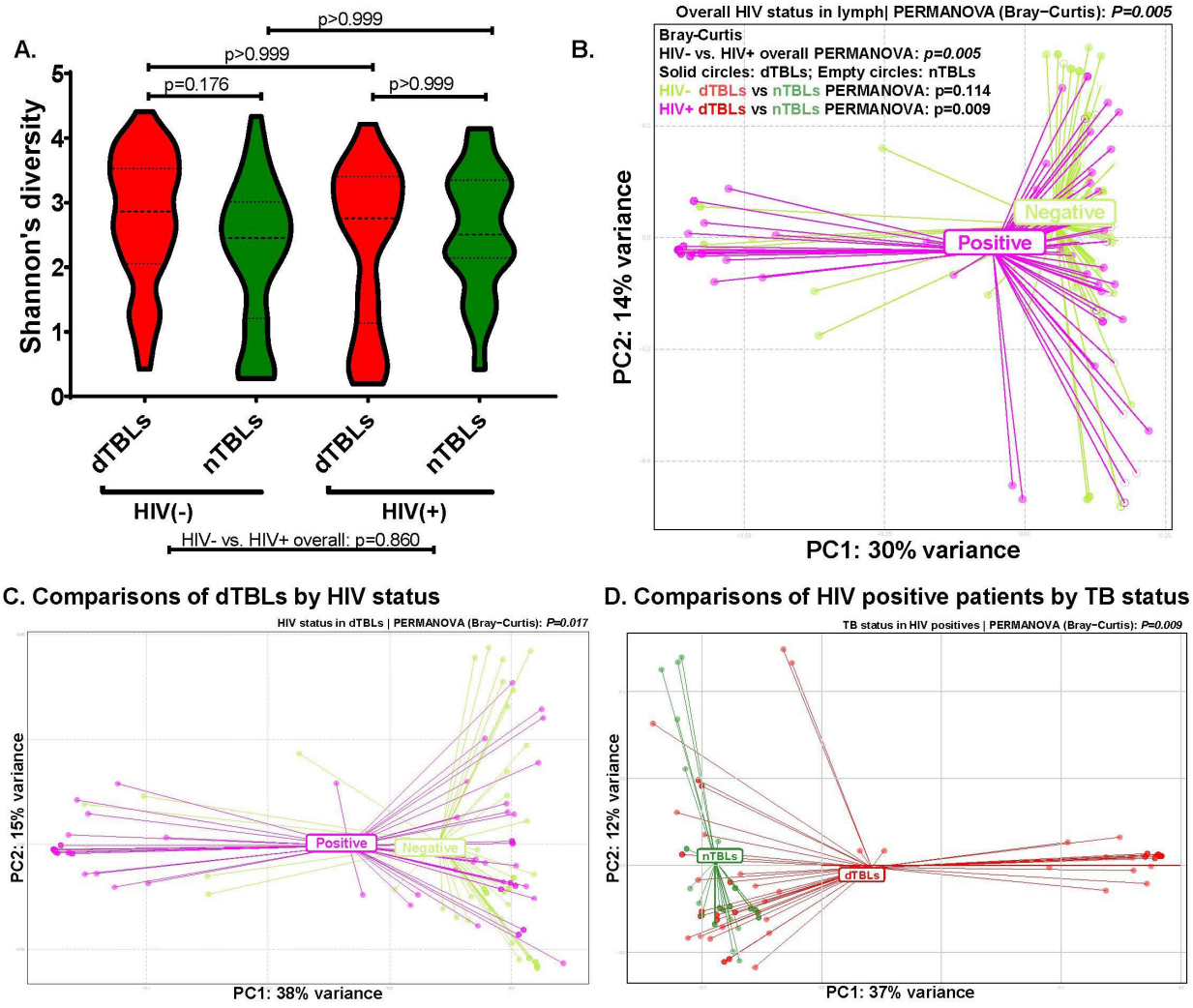


Figure 3 Microbiome differences in HIV-positive dTBLs versus nTBLs but not in HIV-negative dTBLs vs nTBLs. (A) α-Diversity did not differ by HIV or TBL statuses, (B) however, β-diversity differed between HIV-positives and -negatives overall (shaded circles are dTBLs, empty circles are nTBLs). β-diversity differed (C) by HIV status in dTBLs only and (D) by TBL status in HIV-positives only. d-TBLs, definite TBLs; nTBLs, non-TBLs; PERMANOVA, permutational multivariate analysis of variance; TBL, tuberculous lymphadenitis.

Differences by HIV status

HIV is a known risk factor for TB. We assessed its association with the lymph microbiome first in all patients irrespective of TB status and next within dTBLs or nTBLs. Overall, α-diversity did not differ by HIV status (figure 3A), but β-diversity did (figure 3B). β-diversity differences by HIV status persisted within dTBLs (p=0.017, figure 3C) but not nTBLs. In people with the same HIV status, β-diversity differed between dTBLs vs nTBLs only in HIV-positives (p=*0.009*, figure 3D) where dTBLs were *Mycobacterium*-enriched (online supplemental figure S5B). In PERMANOVA analyses, HIV status was only significantly associated with β-diversity in dTBLs and not nTBLs (online supplemental table S2).

Lymphotype identification and their associations with clinical characteristics

We further explored this data using DMM to identify potential clusters in the TBL microbiome. These clusters were termed 'lymphotypes', and we evaluated associations between each lymphotype(s) and patients' clinical characteristics.

Overall: We examined whether all patients could be grouped into distinct lymphotypes; these were termed overall lymphotypes (oLs). Five oLs with differing α-diversities and β-diversities were identified (figure 4A–C, online supplemental table S3), with the *Mycobacterium*-dominated (figure 4D) oL4 showing the least α-diversity. While no taxa were differentially abundant in oL1 versus other oLs (online supplemental figure S6A–C), oL2, oL3 and oL5 were enriched relative to oL4 in *Corynebacterium*, *Prevotella* and *Streptococcus*, respectively (figure 4E–G). The patients in all oLs were associated with distinct clinical characteristics. The majority of nTBLs occurred in highly diverse oLs with a heterogeneous mixtures of taxa; likely reflecting the spectrum of pathologies in people with TBL ruled out. oL1 was associated with characteristics indicative of less severe lymphadenitis (less TB and HIV involvement). In contrast, oL4 was associated with characteristics resembling more severe lymphadenitis (bigger lymph nodes, chylous FNABs, previous TB, HIV (with a smaller proportion of PLHIV on ART, likely to have lower CD4 counts) and TB involvement. Therefore, in summary, oL1 appears to be associated with less severe forms of lymphadenitis, whereas oL4 was associated more severe forms (online supplemental table S4).

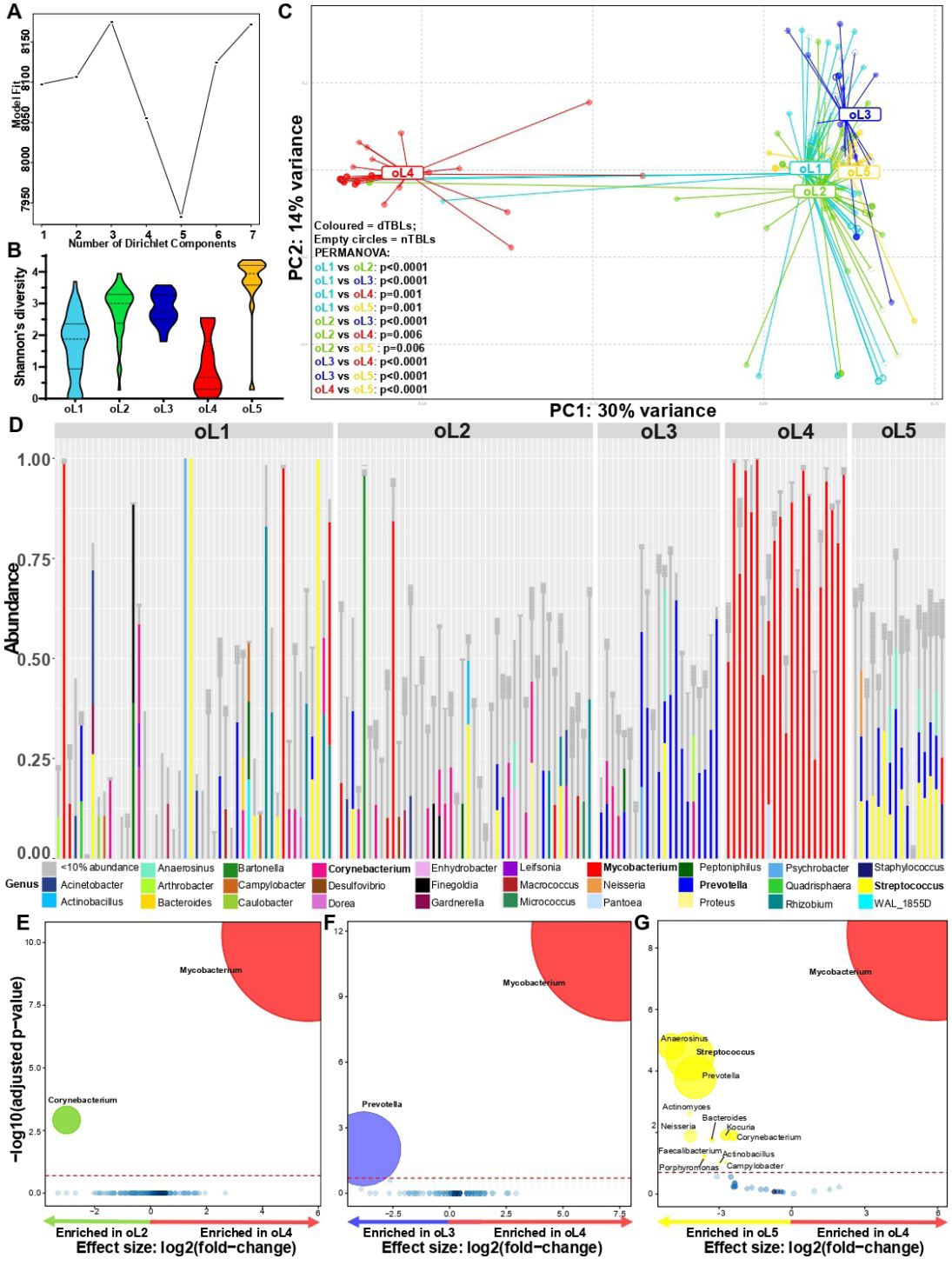


Figure 4 Five overall lymphotypes observed in presumptive TBL. (A) Laplace approximation identified five clusters. (B) OL5 had the highest α-diversity. (C) β-diversity differed between each lymphotype (shaded circles dTBLs, empty circles nTBLs). (D) Stacked bar plots showing OL1 with a heterogeneous mixture of genera, OL2 dominated by *Corynebacterium*, OL3 dominated by *Prevotella*, OL4 dominated by *Mycobacterium*, and OL4 dominated by *Streptococcus*. Bolded taxa represent dominating taxa. (E) *Corynebacterium* was enriched in OL2; (F) *Prevotella* enriched in oL3, (G) *Mycobacterium* enriched in oL4, and *Streptococcus* enriched in OL5. Significantly more discriminatory taxa (bolded) appear closer to the left or right and higher above the threshold (red dotted line, FDR=0.2) as significance increases. Relative taxa abundance is indicated by circle size. dTBLs, definite-TBL; FDR, false discovery rate; nTBLs, non-TBL; oL, overall lymphotype; PERMANOVA, permutational multivariate analysis of variance; TBL, tuberculous lymphadenitis.

Within patients of the same TB status: We then examined whether patients within each TB group could be grouped into distinct lymphotypes. Within dTBLs, three lymphotypes (termed dTBL lymphotypes; dL) with differing β-diversities were identified (figure 5A,B), and dominated by; dL1: *Prevotella* and *Corynebacterium;* dL2: *Prevotella* and *Streptococcus*; and dL3: *Mycobacterium* (figure 5C–F). These dLs were termed *Prevotella-Corynebacterium*,

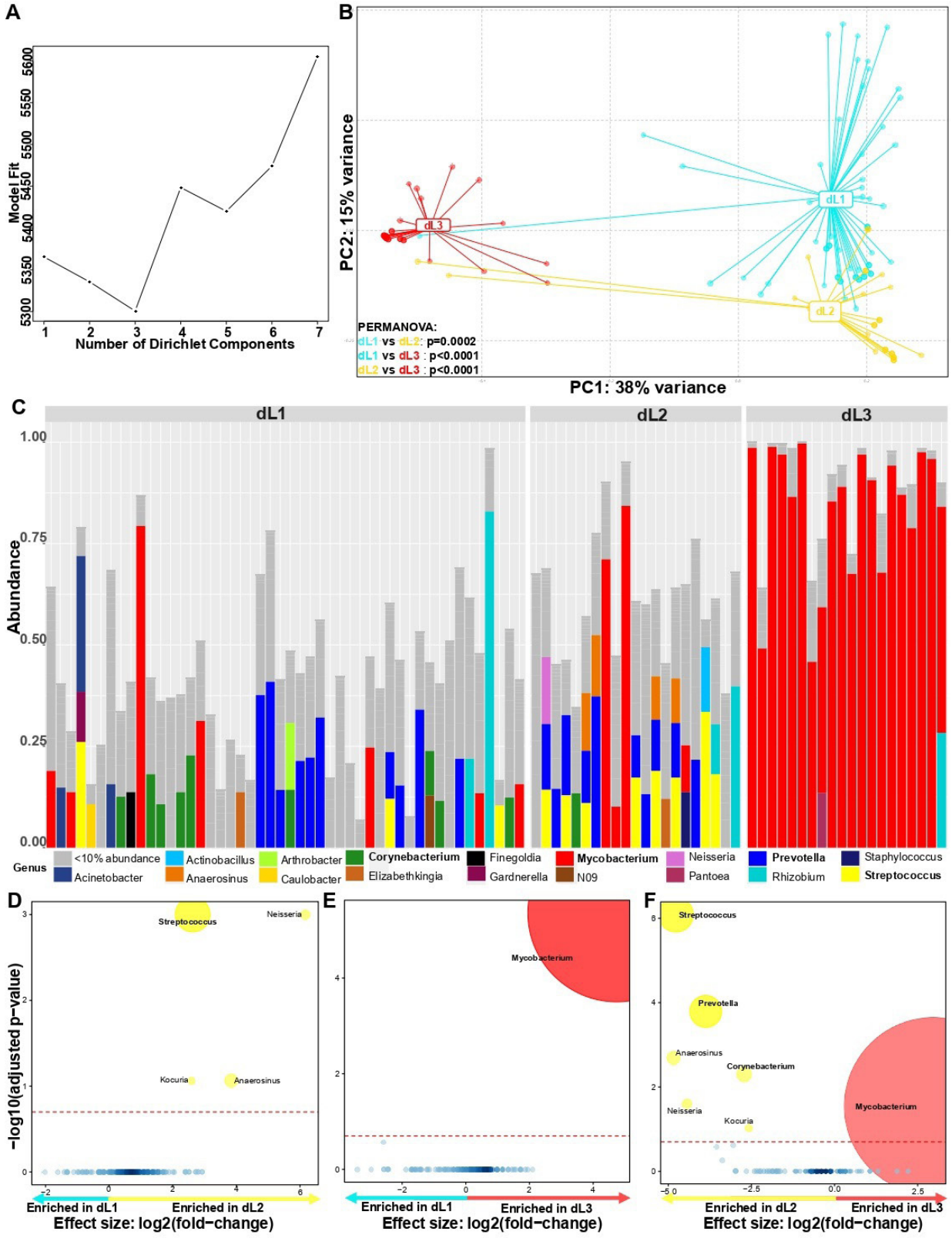


Figure 5 Three dTBL lymphotypes identified in dTBLs. (A) Best model fit based on Laplace approximation identified three clusters within dTBLs. (B) β-diversity differed between lymphotypes. (C) Stacked bar plots showing dL1 comprised of *Mycobacterium* and accompanying heterogenous taxa, dL2 dominated by *Prevotella* and *Streptococus*, and dL3 dominated by *Mycobacterium*. Bolded taxa represent dominating taxa. (D) NO taxa were enriched in dL1, (E) L2 was enriched in *Streptococcus,* (F) and *Mycobaterium* was enriched in dL*3*. Significantly more discriminatory taxa (bolded) appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.2) as significance increases. Relative taxa abundance is indicated by circle size. dL, dTBL lymphotype; dTBL, definite-TBL; FDR, false discovery rate; TBL, tuberculous lymphadenitis.

Prevotella-Streptococcus and *Mycobacterium*, respectively. dL3s were more likely to be HIV-positive compared with dL1s, with larger lymph nodes, compared with dL1s and ddL2s. Lastly, dL2s are more likely to be female than dL1s (online supplemental table S5). Together, these differences suggest dL3 is associated with more severe TBL than other dLs. Within nTBLs, no lymphotypes were identified (online supplemental figure S7).

Predictive metagenome profiling shows increased SCFA metabolism

We further predicted the bacterial metagenome content and made functional inferences of the microbiome using the PICRUSt algorithm. Differences among pathways between groups were evaluated and visualised by *DESeq2* analysis. In dTBLs, 'fatty acid metabolism', 'benzoate degradation', 'propanoate metabolism' and 'butanoate metabolism' were enriched, suggesting increased SCFA production (figure 6). These SCFA-related pathways were enriched in PLHIV overall and, within dTBLs (figure 7A,B).

In addition, when comparing inferred pathways in the 5 oLs, a similar core of pathways was enriched in oL4. In contrast, versus oL4, oL1 was enriched in 'epithelial cell signalling in *Helicobacter pylori* infection', oL2 and oL5 were enriched in 'carbohydrate digestion and absorption', and oL3 was enriched in 'dioxin degradation' (online supplemental figure S9A–H). When comparing the three dLs, *Mycobacterium*-dominated oL3 was, compared with each other dLs, enriched in the similar core

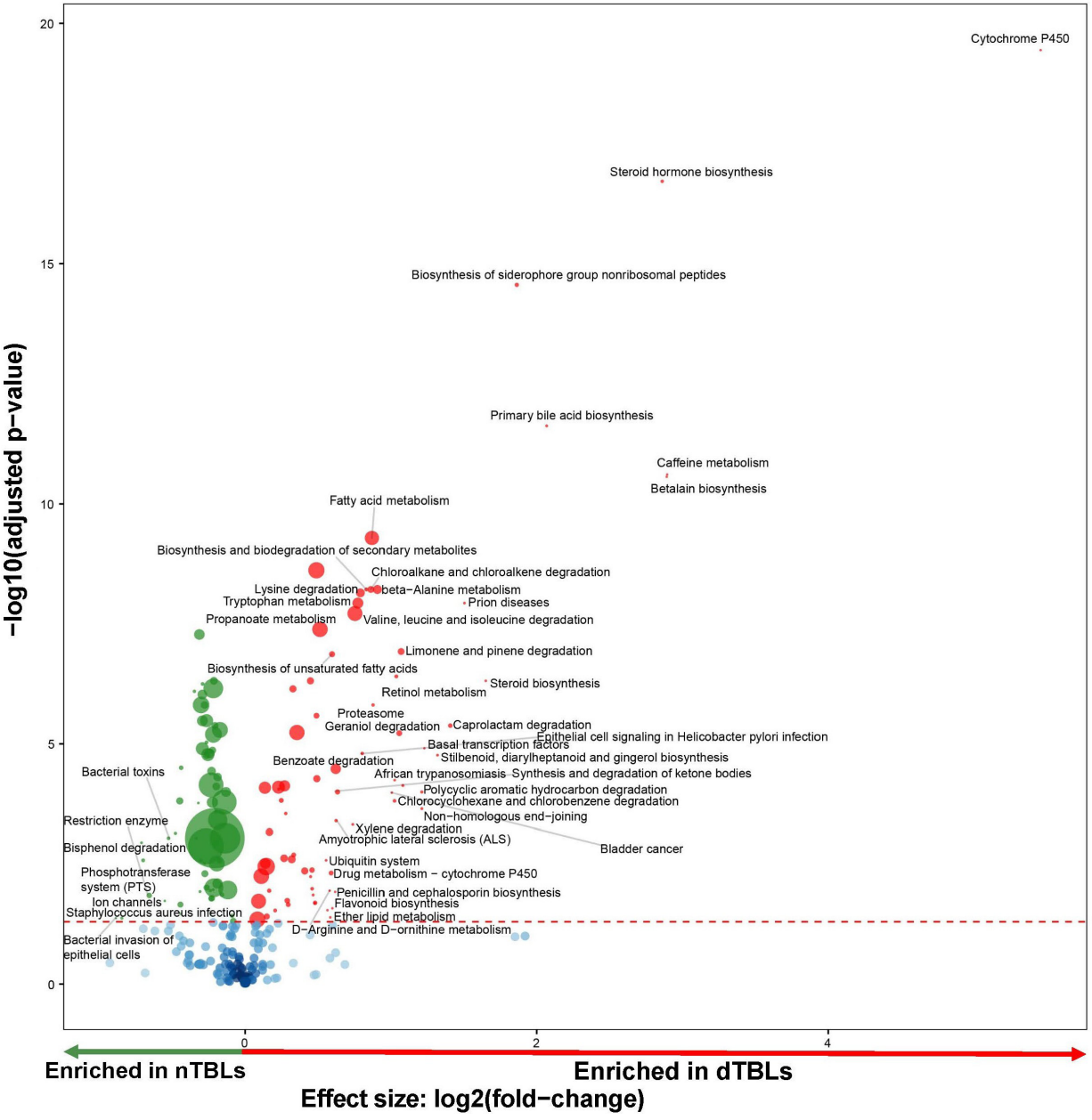


Figure 6 Enriched microbial capacity for SCFA pathways in dTBLs vs nTBLs. Volcano plot depicting differentially abundant microbial pathways in dTBLs vs nTBLs inferred by PICRUSt2. key pathways of interest are bolded including aminobenzoate degradation, benzoate degradation and propanoate degradation. Significantly more discriminatory pathways appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.05) as significance increases. Relative gene abundance is indicated by circle size. dTBLs, definite-TBLs; FDR, false discovery rate; nTBLs, non-TBL; SCFA, short-chain fatty acids.

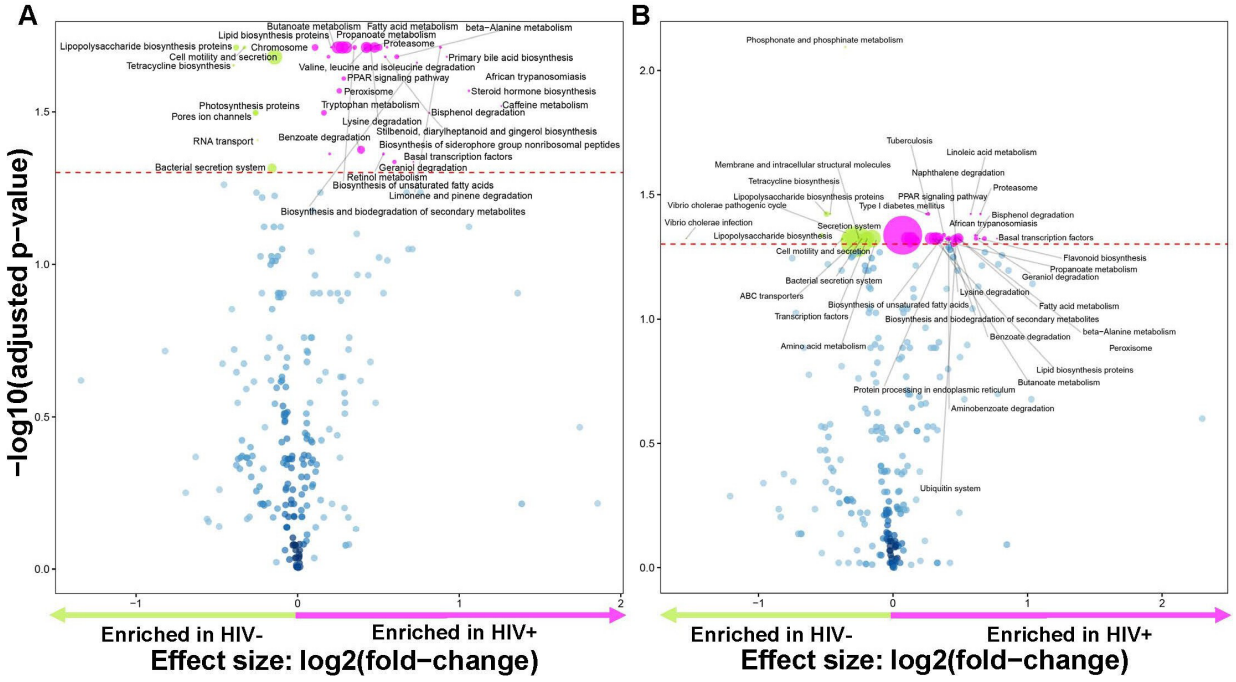


Figure 7 Predicted metagenome function reveals increased capacity for SCFA production in HIV-positive versus HIV-negative patients overall, and in dTBLs. Volcano plot depicting functional pathways differing between (A) HIV-positive and HIV-negative patients with presumptive TBL and (B) in dTBLs. Key pathways of interest include butanoate metabolism, propanoate metabolism and benzoate degradation. Significantly more discriminatory pathways appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.05). Relative pathway abundance is indicated by circle size. dTBLs, definite-TBL; FDR, false discovery rate; SCFA, short-chain fatty acids.

pathways as the *Mycobacterium*-dominated oL4 in all patients (figure 8C; online supplemental figure S10). These results show that pathways involved in fatty acid-related, amino acid-related and SCFA-related inferred microbial pathways were significantly enriched in dTBLs and *Mycobacterium* lymphotypes (oL4 and dL3).

DISCUSSION

We characterised the local microbial environment in patients with lymphadenitis undergoing investigation for TB in an HIV-endemic setting. Our key findings are: (1) lymphatic microbial communities in dTBLs clustered into three distinct 'lymphotypes' we termed '*Prevotella-Corynebacterium*', '*Prevotella-Streptococcus*' and '*Mycobacterium*', (2) the *Mycobacterium* dTBL lymphotype was associated with HIV-positivity and other clinical features characteristic of severe lymphadenitis and (3) dTBLs relative to nTBLs were functionally enriched in fatty acid-related, amino acid-related and SCFA-related microbial metabolic pathways with known immunomodulatory effects (the *Mycobacterium* lymphotype was most enriched in these pathways than other dTBL lymphotypes). Finally, (4) dTBLs without *Mycobacterium* reads and nTBLs with *Mycobacterium* reads were identified. These data show TBL at the site-of-disease is not microbially homogenous and that distinct clusters of microbial communities exist associated with different clinical characteristics. The long-term significance and importance of these lymphotypes requires prospective evaluation.

We identified three lymphotypes within dTBLs termed '*Prevotella-Corynebacterium*', '*Prevotella-Streptococcus*' and '*Mycobacterium*', distinguished by different relative abundances of these taxa (*Prevotella* co-occurred in the first two lymphotypes). These individual taxa are enriched in respiratory secretions from pulmonary TB cases.[31 32] Furthermore, within dTBLs, *Streptococcus* is associated with low BMI and extent of lung damage.[32] *Prevotella* in BALF also positively correlates with SCFA concentrations and independently predicts incident TB in people without co-prevalent TB.[13] Compared with the other dTBL lymphotypes, '*Mycobacterium*' was associated with severe disease and most frequently occurred in PLHIV, agreeing with diagnostics studies that show stronger baseline mycobacterial PCR test readouts predict long term clinical outcomes in pulmonary[33] and extrapulmonary TB.[34] Together, these data show distinct lymphotypes are associated with different clinical characteristics and suggests that patients with the most severe *Mycobacterium*-dominated lymphotype may initially progress through different site-of-disease microbial states characterised by *Corynebacterium*, *Streptococcus* and/or *Prevotella* domination. Studies with longitudinal follow-up and repeat sampling are required to examine whether these lymphotypes have potential for clinical staging.

Importantly, *Corynebacterium* and *Streptococcus* often dominated in dTBL patients. Members of both taxa are causative agents of lymphadenitis and, even though these patients have TBL confirmed via conventional diagnostics, *Corynebacterium* and *Streptococcus* may therefore cocontribute to pathology and symptoms.[35–37] Coincidently, these taxa fall within the antimicrobial spectrum of first-line TB treatment,[16] meaning that this regimen may, in part, cure lymphadenitis by killing *Corynebacterium* and *Streptococcus* in addition to *Mycobacterium*.

Microbial pathways predicted to be most enriched in dTBLs involved fatty acid, amino acid and SCFAs (benzoate, propanoate) metabolism; all of which are associated with pulmonary TB disease compared with sick patients without TB.[38 39] SCFAs in particular suppress immune pathways involved in IFN-γ and IL-17A production and, ex vivo, limit macrophage-mediated kill of *Mtb*. SCFA concentrations hence predict incident TB in

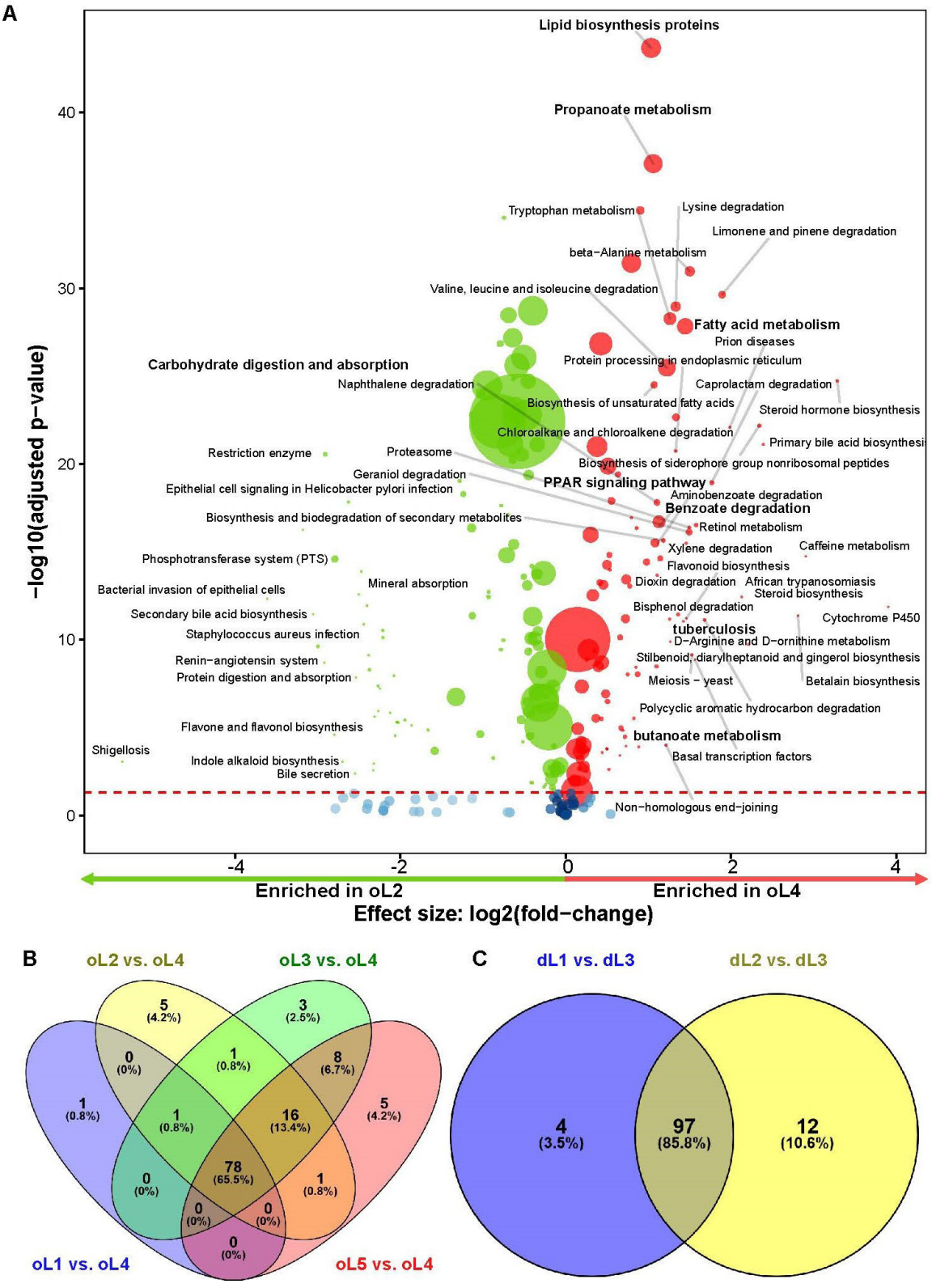


Figure 8 Differential microbial pathways between lymphotypes showing similar core pathways enriched in the *Mycobacterium*-dominated lymphotype. (A) Volcano plot showing differentially abundant microbial pathways inferred by PICRUSt2 in oL2 vs oL4 representing pathways enriched in oL4 compared with every other oL in all patients (overall including dTBLs and nTBLs). Significantly more discriminatory pathways appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.05) as significance increases. Relative pathway abundance is indicated by circle size. (B) 65.5% of all inferred pathways enriched in oL4 compared with each other oLS were common, while (C) 85.8% were common in dL3 compared with each other dTBL lymphotypes.Differentially enriched pathways common in all comparisons with the *Mycobacterium* dominant lymphotype included pathways involving lipid biosynthesis, fatty acids and SCFA metabolism, that is, lipid biosynthesis proteins, propanoate metabolism, benzoate degradation, and valine, leucine and isoleucine degradation. dL, dTBL lymphotype; dTBLs: definite-TBLs; FDR, false discovery rate; nTBLs: non-TBL; oL: overall lymphotype; SCFA: short-chain fatty acid.

patients.[13] Our research therefore suggests that the inflammation associated with lymphadenopathy is in part caused by the presence of microbes including but not limited to *Mycobacterium* that are able to produce SCFAs that interfere with these immunological pathways; revealing potentially new therapeutic targets to reduce lymphadenopathy.

We detected *Mtb* DNA in nTBLs. These reads could be from subclinical infection, previous TB exposure or disease, where the DNA was transported to the lymph node. *Mtb* DNA has been found in the lymph nodes of healthy individuals and primates exposed to TB, where the sites are hypothesised to serve as a *Mtb* growth and persistence niche.[6] dTBLs without *Mycobacterium* reads were also documented, however, 16S rRNA sequencing has known suboptimal sensitivity for *Mycobacterium*, in part due to low 16s RNA gene copy number.[40]

Our study has strengths and limitations. Patients were sampled once, as close as possible to treatment initiation; animal models might permit repeat invasive sampling especially if treatment is withheld. The programmatic context enabled large numbers of patients to be recruited, however, detailed long-term follow-up, which could include imaging of lymph nodes and more detailed measurements of differential responses to treatment, was not possible. We did not perform any viability tests, and since 16S gene sequencing is DNA based, the DNA may have originated from live, dead or nonculturable bacteria. Future studies could use meta-transcriptomics or culturomics to investigate this. We also used an FDR-adjusted p value threshold of 0.2 to identify differentially abundant taxa because this study was designed to be hypothesis generating and lower thresholds did not generate such taxa. Furthermore, the use of PICRUSt to infer potential function from 16S rRNA gene sequencing is a limitation. Follow-up studies using shotgun metagenomics, are necessary for inferring biological function and can more comprehensively describe the microbiota beyond bacteria. Our study was designed to describe the site-of-disease microbiome in TBL in a setting with a high burden of TB and HIV. Further research in different settings and populations is needed to validate our findings, especially those findings pertaining to microbial community clustering and the relationship between individual clusters and clinical characteristics.

In conclusion, we show dTBL patients have a distinct microbiome at the site of disease, characterised by three lymphotypes (*Mycobacterium*, *Prevotella-Corynebacterium*, *Prevotella-Streptococcus*). This dysbiosis of the lymphatic microbiome likely contributes to pathophysiology, including inflammatory state and clinical severity, which itself may reflect the chronicity of TB disease. TBL does therefore not appear to be a microbially homogenesis disease, and this reveals potentially new diagnosis, therapeutic and prognostic targets.

Author affiliations
[1]DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa
[2]African Microbiome Institute, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa
[3]Division of Pulmonary and Critical Care Medicine, New York University Grossman School of Medicine, NYU Langone Health, New York, NY, USA
[4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[5]National Health Laboratory Service, Tygerberg Hospital, Cape Town, Western Cape, South Africa
[6]Division of Medical Microbiology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa
[7]Division Anatomical Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa

Correction notice This article has been corrected since it was first published. The open access licence has been updated to CC BY.

Acknowledgements The authors thank study participants and Tygerberg Hospital FNA clinic staff especially Sr Cupido. Additionally, we thank CLIME research group staff, especially Sr Ruth Wilson and Roxanne Higgit. GRN acknowledges funding from L'Oréal-UNESCO For Women in Science Sub-Saharan Africa Young Talents Award, and the International Rising Talents Award. The content is the sole responsibility of the authors and does not necessarily represent the official views of the funders. Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: hpc.uct.ac.za.

Contributors GRN, CCN and GT contributed to conceptualisation and design of the study and supervised the study, funding acquisition, data collection, and wrote the manuscript. GRN, CCN, IS, BW, IS, JCC, LS and GT contributed to data and statistical analysis, and figures. All authors contributed to interpretation of data and editing of the manuscript. As the study guarantor, GT is responsible for the overall content of this manuscript.

Funding This work and authors were supported by the European & Developing Countries Clinical Trials Partnership (EDCTP; project numbers SF1041, TMA2017CDF-1914-MOSAIC and TMA2019CDF-2738-ESKAPE-TB), National Research Foundation (NRF), the South African Medical Research Council (SAMRC), the Harry Crossley Foundation and Stellenbosch University Faculty of Health Sciences, and the National Institutes of Health under award numbers (R01AI136894; U01AI152087; U54EB027049; D43TW010350; K43TW012302).

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Health Research and Ethical Committee of Stellenbosch University (N16/04/050), Tygerberg Hospital (Project ID:4134),Western Cape Department of Health (WC_2016RP15_762). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data are available on reasonable request. Deidentified patient data, the study protocol, informed consent and datasets generated in this study may be requested from the corresponding author. Sequencing data reported in this paper will be released and made publicly available for sharing on publication (NCBI Bioproject PRJNA738676).

ORCID iD
Grant Theron http://orcid.org/0000-0002-9216-2415

REFERENCES

1 WHO. Global tuberculosis report 2020. Geneva, Switzerland. World Health Organization; 2020.
2 Gupta RK, Lawn SD, Bekker L-G, *et al*. Impact of human immunodeficiency virus and CD4 count on tuberculosis diagnosis: analysis of city-wide data from Cape town, South Africa. *Int J Tuberc Lung Dis* 2013;17:1014–22.
3 Mohapatra PR, Janmeja AK. Tuberculous lymphadenitis. *J Assoc Physicians India* 2009;57:585–90.
4 Mekonnen D, Derbie A, Abeje A, *et al*. Epidemiology of tuberculous lymphadenitis in Africa: a systematic review and meta-analysis. *PLoS One* 2019;14:e0215647.
5 Maji A, Misra R, Kumar Mondal A, Mondal AK, *et al*. Expression profiling of lymph nodes in tuberculosis patients reveal inflammatory milieu at site of infection. *Sci Rep* 2015;5:1–10.
6 Ganchua SKC, White AG, Klein EC, *et al*. Lymph nodes-The neglected battlefield in tuberculosis. *PLoS Pathog* 2020;16:e1008632.

7 Martin CJ, Cadena AM, Leung VW, *et al*. Digitally Barcoding *Mycobacterium tuberculosis* Reveals *In Vivo* Infection Dynamics in the Macaque Model of Tuberculosis. *mBio* 2017;8. doi:10.1128/mBio.00312-17. [Epub ahead of print: 09 May 2017].
8 Yik Wang C, Wang CY. An experimental study of latent tuberculosis. *The Lancet* 1916;188:417–9.
9 Ganchua SKC, Cadena AM, Maiello P, *et al*. Lymph nodes are sites of prolonged bacterial persistence during Mycobacterium tuberculosis infection in macaques. *PLoS Pathog* 2018;14:e1007337.
10 Behr MA, Waters WR. Is tuberculosis a lymphatic disease with a pulmonary portal? *Lancet Infect Dis* 2014;14:250–5.
11 Esmail H, Lai RP, Lesosky M, *et al*. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[^{18}F]fluoro-D-glucose positron emission and computed tomography. *Nat Med* 2016;22:1090–3.
12 Blacher E, Levy M, Tatirovsky E, *et al*. Microbiome-modulated metabolites at the interface of host immunity. *J Immunol* 2017;198:572–80.
13 Segal LN, Clemente JC, Li Y, *et al*. Anaerobic bacterial fermentation products increase tuberculosis risk in Antiretroviral-Drug-Treated HIV patients. *Cell Host Microbe* 2017;21:530–7.
14 Kiernan MG, Coffey JC, McDermott K, *et al*. The human mesenteric lymph node microbiome differentiates between Crohn's disease and ulcerative colitis. *Journal of Crohn's and Colitis* 2019;13:58–66.
15 O'Brien CL, Pavli P, Gordon DM, *et al*. Detection of bacterial DNA in lymph nodes of Crohn's disease patients using high throughput sequencing. *Gut* 2014;63:1596–606.
16 Naidoo CC, Nyawo GR, Wu BG, *et al*. The microbiome and tuberculosis: state of the art, potential applications, and defining the clinical research agenda. *Lancet Respir Med* 2019;7:892–906.
17 Vázquez-Pérez JA, Carrillo CO, Iñiguez-García MA, *et al*. Alveolar microbiota profile in patients with human pulmonary tuberculosis and interstitial pneumonia. *Microb Pathog* 2020;139:103851.
18 Zhou Y, Lin F, Cui Z, *et al*. Correlation between either Cupriavidus or Porphyromonas and primary pulmonary tuberculosis found by analysing the microbiota in patients' bronchoalveolar lavage fluid. *PLoS One* 2015;10:e0124194.
19 Minnies S, Reeve BWP, Rockman L, *et al*. Xpert MTB/RIF ultra is highly sensitive for the diagnosis of tuberculosis lymphadenitis in a High-HIV setting. *J Clin Microbiol* 2021;59:e01316–21.
20 Bolyen E, Rideout JR, Dillon MR, *et al*. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;37:852–7.
21 Callahan BJ, McMurdie PJ, Rosen MJ, *et al*. DADA2: high-resolution sample inference from illumina amplicon data. *Nat Methods* 2016;13:581–3.
22 McDonald D, Price MN, Goodrich J, *et al*. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *Isme J* 2012;6:610–8.
23 McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;8:e61217.
24 Dixon P. Vegan, a package of R functions for community ecology. *J Veg Sci* 2003;14:927–30.
25 Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* 2012;7:e30126.
26 Douglas GM, Maffei VJ, Zaneveld J. PICRUSt2: an improved and customizable approach for metagenome inference. *BioRxiv* 2020;672295.
27 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 2014;15:1–21.
28 Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 2003;19:368–75.
29 SSTS. *Stata: data analysis and statistical software. Stata statistical software: release 16*. College Station, TX, 2022.
30 Estaki M, Jiang L, Bokulich NA, *et al*. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Curr Protoc Bioinformatics* 2020;70:e100.
31 Krishna P, Jain A, Bisen PS. Microbiome diversity in the sputum of patients with pulmonary tuberculosis. *Eur J Clin Microbiol Infect Dis* 2016;35:1205–10.
32 Ticlla MR, Hella J, Hiza H, *et al*. The sputum microbiome in pulmonary tuberculosis and its association with disease manifestations: a cross-sectional study. *Front Microbiol* 2021;12:633396.
33 Pires MdeM, Pereira GR, Barbosa MS, *et al*. Association of Xpert MTB/RIF cycle threshold values with tuberculosis treatment outcomes. *Lung* 2020;198:985–9.
34 Martyn EM, Bangdiwala AS, Kagimu E, *et al*. Cerebrospinal fluid bacillary load by Xpert MTB/RIF ultra polymerase chain reaction cycle threshold value predicts 2-week mortality in human immunodeficiency virus-associated tuberculous meningitis. *Clin Infect Dis* 2021;73:e3505–10.
35 Peel MM, Palmer GG, Stacpoole AM, *et al*. Human lymphadenitis due to Corynebacterium pseudotuberculosis: report of ten cases from Australia and review. *Clin Infect Dis* 1997;24:185–91.
36 Barroso LF, Pegram SP. Clinical manifestations, diagnosis, and treatment of diphtheria. In: *UpToDate*. Post TW ed. Waltham, MA: UpToDate, 2022.
37 Denis Spelman M, Baddour LM. Cellulitis and skin abscess: Epidemiology, microbiology, clinical manifestations, and diagnosis. In: *UpToDate*. Post TW ed. Waltham, MA: UpToDate, 2021.
38 Naidoo CC, Nyawo GR, Sulaiman I, *et al*. Anaerobe-enriched gut microbiota predicts pro-inflammatory responses in pulmonary tuberculosis. *EBioMedicine* 2021;67:103374.
39 Somboro AM, Diallo D, Holl JL, *et al*. The role of the microbiome in inflammation during tuberculosis. *EBioMedicine* 2021;68:103435.
40 Sulaiman I, Wu BG, Li Y, *et al*. Evaluation of the airway microbiome in nontuberculous mycobacteria disease. *Eur Respir J* 2018;52. doi:10.1183/13993003.00810-2018. [Epub ahead of print: 25 Oct 2018].